A STUDY OF THE RELATIONSHIPS BETWEEN INFORMAL SECOND LANGUAGE CONTACT, VOCABULARY-RELATED STRATEGIC BEHAVIOUR AND VOCABULARY GAIN IN A STUDY ABROAD CONTEXT

Jessica G. Briggs

KELLOGG COLLEGE

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN EDUCATION

TRINITY TERM, 2014

VOLUME I

TABLE OF CONTENTS

ACKNOWLEGEMENTS

With sincerest thanks to:

Professor Ernesto Macaro, for his guidance and support and for being human as well as superhuman.

The Economic and Social Research Council, for their financial support.

Dr Catherine Walter, for invaluable work experience and an inspirational role model.

Dr Bob Vanderplank, Dr Catherine Walter, Prof Vicki Murphy, Dr Xin Wang and Dr Peter Yongqi Gu for reading and giving feedback on this study at various stages of its development.

Dr James Hall, for helping me to love multiple regression analysis.

The language learners who participated in this study, for generously giving their time and for being a thoroughly interesting group of people to spend time with.

Dr Sara Smith and Dr Charlotte Clancy, for always being a few steps ahead of me and letting me know what to expect.

Howard and Sue Briggs, for their unwavering love, support and belief in me.

Sue Hicks, Kojo Minta and Omar Briggs, for having been and being still in my heart.

George Baffoe-Djan, for the constancy, the Excel formulae, the hugs and the dances. I couldn't have done it without you, and I wouldn't have wanted to.

ABSTRACT

This thesis reports on a longitudinal, mixed-methods study of the relationships between informal (i.e. out-of-class) second language (L2) contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context. The study addressed three main gaps in knowledge that arose from analysis of the literature: (1) the evidence of informal L2 contact was largely unreliable, ungeneralisable, or both; (2) the evidence of vocabulary-related strategic behaviour in informal L2 contact was neither context nor task specific; and (3) there was no evidence of the interplay between informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context.

The sample (n=241) were adults undertaking a study abroad experience (SAE) in England, who comprised a range of nationalities and first language backgrounds and for whom the majority of the SAE was spent outside of the classroom. A vocabulary test was administered at the beginning and end of the SAE. A questionnaire was administered during the SAE to determine the most highly identified with informal L2 contact scenarios and out-of-class vocabulary-related strategies. Subsequently, an innovative research tool comprising computer-based simulations of the most identified with scenarios was developed and used as the stimulus in semi-structured interviews to capture task and/or context-specific vocabulary-related strategic behaviour. Analysis grouped participants by length of stay and location.

The most highly identified with informal L2 contact scenarios involved participants seeking information from external sources, such as interlocutors, posters or websites. The vocabulary-related strategies most highly identified with by the sample pertained to the *use* of a newly encountered lexical item; that is, they were strategies in which the learner used or prepared to use a lexical item that they had decided to engage with strategically. The strategic behaviour manifested in response to the simulation tool (the 'OWLS') provided strong evidence in support of the fundamental considerations of task, context and intention in strategy-based research. Regression analysis revealed that informal L2 contact scenarios that were less strategically prohibitive and strategies that were less context-dependent were predictors of vocabulary gain. The pedagogical implications of these findings are far-reaching in terms of preparing L2 learners for informal contact on a SAE and guiding their manipulation of that contact for maximum linguistic gain.

LIST OF TABLES AND FIGURES

TABLES

FIGURES

LIST OF ABBREVIATIONS

ACTFL	American Council on the Teaching of Foreign Languages
AH	At home
ANOVA	Analysis of variance
AWL	Academic Word List
BAAL	British Association for Applied Linguistics
BNC	British National Corpus
CAQDAS	Computer-assisted qualitative data analysis software
CEFR	Common European Framework of Reference
CELTA	Certificate in English Language Teaching to Adults
CFA	Confirmatory factor analysis
CUREC	Central University Research Ethics Committee
EFA	Exploratory factor analysis
EFL	English as a foreign language
ELT	English language teaching
EMI	English medium instruction
ESL	English as a second language
ESOL	English to speakers of other languages
DELTA	Diploma in English Language Teaching to Adults
GSL	General Service List
HSD	Honestly significant difference
IELTS	International English Language Testing System
L1	First language
L2	Second language
L3	Third language
LCP	Language Contact Profile
M	Mean
MMR	Mixed methods research
OPI	Oral Proficiency Interview
OWLS	Opportunities With Language Simulator
SAE	Study abroad experience
SASIQ	Study Abroad Social Network Questionnaire
SD	Standard deviation
SE	Standard error
SILL	Strategies Inventory for Language Learning
SLA	Second language acquisition
SPSS	Statistical Package for the Social Sciences
TM	Trimmed mean
TOEFL	Test of English as a Foreign Language
PV	Phrasal verb
Q	Questionnaire
QUAN	Quantitative

QUAL	Qualitative
RQ	Research question
UK	United Kingdom
VT	Vocabulary test

CHAPTER I

INTRODUCTION

Context of learning has been a key issue of research and discussion within the second language acquisition field of study, yet of the three main contexts of second language (L2) learning (the English as a foreign language classroom, the immersion setting and study abroad experience), study abroad experience (SAE) has received comparatively little attention. Approximately 700,000 individuals per annum travel to the UK to study English (British Council, 2014) and research suggests that SAE has a significant impact on second language acquisition (e.g. Dewey, 2008). Most learners' time during a SAE is spent outside the classroom (c.f. section 3.3), yet research has so far failed to address in sufficient depth the specific nature of the contact learners have with the L2 during this time, nor the behaviours they manifest in response. Instead, studies have tended toward quantifying the amount of informal contact that learners encounter. This is not necessarily surprising given the complexities inherent in researching language acquisition which occurs outside of a controllable setting, yet there is a strong case for determining in fine detail the out-of-class scenarios which study abroad learners most identify with and how they respond to unfamiliar vocabulary in those scenarios in order to determine how vocabulary gain can be maximised during SAE.

This study identifies and examines the relationships between informal L2 contact, the vocabulary-related strategic behaviour manifested by learners of English in that contact, and L2 vocabulary gain in a study abroad context. It is intended that the findings here reported will provide empirically-grounded guidance for learners, teachers and the study

abroad industry with respect to what is currently a largely uninformed but potentially fundamental aspect of study abroad: informal, learner-managed contact with the L2.

1.1 STUDY ABROAD EXPERIENCE

Study abroad is unique as a language learning context because of the constant availability of L2 exposure. In the UK for example, the minimum instruction required for a General Student Visa is fifteen hours per week and students often cohabit with people who do not speak their L1 (e.g. with a host family belonging to the target language community). This uniqueness has led to a number of studies investigating the difference between SAE and other language learning settings regarding their impact on proficiency gain (e.g. Collentine, 2004). However, there is a dearth of research into the informal L2 contact that study abroad learners encounter. Neither have comparisons been made between different L2 contact scenarios in terms of their potential impact on second language acquisition: Are some scenarios more beneficial than others to engage in? Which are likely to beget meaningful language processing? Are we able to hypothesise about the strategic behaviours learners will manifest? It is questions such as these that remain largely unanswered yet have the potential dramatically to impact upon the linguistic outcomes of SAE.

Many language learners leave employment or delay starting higher education to undertake a SAE, believing a stay in the target language community to be the *sine qua non* of second language mastery. Tuition costs can be high, and the cost of living higher than in many learners' home countries. Meara (1994) states that despite the investment required to undertake SAE, research has yet to establish its value. Moreover, 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 a SAE promises (Tanaka, 2004; Woodman, 1998). However, the rhetoric purveyed online by the study abroad industry fails to reference the potential drawbacks and limitations to a sojourn abroad and instead paints a wholly rosy picture of the linguistic benefits it promises to those who have the requisite financial and temporal means. For example, Eurolingua (2014) 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 online about the L2 contact that learners will experience on a SAE:

"*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, 2014).

Therefore, the focus in this study on the study abroad context is derived not only from gaps in the collective knowledge about study abroad as highlighted by the analysis of the literature (and summarised in section 2.4.4), but also from a desire to limit the extent to which study abroad learners are sold false promises and unrealistic expectations about SAE.

1.2 VOCABULARY KNOWLEDGE

"When students travel, they don't carry grammar books, they carry dictionaries" (Krashen, 1987). The decision to focus on vocabulary gain in this study derives from the importance of vocabulary to all the modalities of the second language; indeed, learners themselves consider vocabulary knowledge a significant contributor to both their receptive and productive ability (Nation, 1990), and Wilkins (1972) states that although much can be communicated without grammar, nothing is communicated without vocabulary.

Research has provided strong evidence as to the primacy of the role of vocabulary in second language competence. Vocabulary knowledge has variously 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, vocabulary size may in fact impact on other areas of linguistic competence, such as grammatical competence (Myles, Hooper & Mitchell, 1998; Bardovi-Harlig, 2002) and phonological control (Bundgaard-Nielsen, Best & Tyler, 2011a, 2011b).

1.3 STRATEGIC BEHAVIOUR

The benefit of individual or combinations of strategies on second language acquisition is said to depend in part on the context in which they are used (Schmitt & McCarthy, 1997) and therefore an understanding of out-of-class strategy use which correlates positively with vocabulary gain during a SAE needs to be established. With this knowledge, learners and educators will be better equipped to exploit the myriad of informal L2 contact scenarios that SAE has the potential to provide.

Over thirty years ago, Bialystok (1983) noted that disparity existed among scholars with regard to what constitutes a language learning strategy, and some twenty-five years later the situation was little changed (Grenfell & Macaro, 2007). Some researchers' definitions of language learning strategies (Oxford, 1990) suggest that they are behavioural and therefore observable, whereas others (Weinstein & Mayer, 1986) extend their definition to encompass unobservable mental action. Some (Seliger, 1983; Ellis, 1985) challenge the notion that any strategic behaviour is observable, believing that behind a physical action lies a cognitive process that is almost wholly unobservable. Stevick (1990) posits the 'size-abstractness dilemma' to argue that the phenomena to which strategies pertain are often very different in size and varied in abstractness. Researchers (Sparks & Ganschow, 1993; Stevick, 1990) have also argued that using observable action to define language learning strategies is too global: one physical action could derive from any number of different mental intentions.

Given, then, that strategies can be unobservable, to research strategies in settings which by their very nature cannot be researcher-controlled or in real-time directly observed is an undertaking which has not, until now, been undertaken. For use in this study an innovative research tool, the Opportunities With Language Simulator (OWLS), was designed and employed to capture in fine detail the vocabulary-related strategies that might be used by adult learners of English on a SAE when exposed to unfamiliar vocabulary outside of the classroom. Comprising computer-based simulations of the most highly identified with informal L2 contact scenarios, the tool was designed for use as a stimulus in semi-structured interviews in order to capture the task and/or context-specific strategic behaviour that learners manifest. Use of the OWLS in a previous study (Briggs, 2014) revealed the power of the tool to illuminate the relationship between context and action by generating fine-grained introspective data on strategic behaviour, the like of which is

unmatched by other existing research instruments. Moreover, the OWLS demonstrates that context, and particularly contextual intention (that is, the non-linguistic aim of a learner in a given informal L2 contact scenario) has a significant impact on strategic behaviour in response to unknown vocabulary encountered in informal settings.

1.4 FRAMEWORKS OF REFERENCE

Here follows a discussion of the conceptual and theoretical frameworks underpinning the study respectively.

1.4.1 CONCEPTUAL FRAMEWORK

The study here reported employs a mixed methods design. Mixed methods research (MMR) describes the combination in a single study of techniques commonly associated with either quantitative or qualitative research; the theoretical drive of the study derives from one of these two research traditions (which acts as the principal method), and the strategies imported from the other research tradition supplement the data generated by the principal method (Morse, 2003). MMR emerged from the crude but long-perpetuated dichotomy of the paradigm 'wars' of the quantitative (QUAN) and qualitative (QUAL) research traditions (Morgan, 2007; Tashakkori & Teddlie, 2003), offering a pragmatic alternative to dualism and competing epistemologies.

As Dörnyei (2007) states, quantitative research enjoys high status because it offers methodical and exact treatment of data using sophisticated statistical instruments that are relatively quick and simple to use, yet the tendency of quantitative research to generalise is criticised by some as being overly simplistic and inconsiderate of individuals and context. The strength of qualitative research is in the depth of information it provides, making it well suited to exploring idiosyncrasies in individuals' behaviour (Myers, 2000). Rather

than aiming to generalise, qualitative research offers a range of possible explanations for phenomena and is inclusive in that all information can be regarded as data (Dörnyei, 2007). However, the process of reaching said explanations is highly labour intensive, and the researcher is heavily relied upon to interpret the data without bias or subjectivity. The strength of a mixed methods design therefore is the combination of rigorous, objective, numeric findings with detailed and individualised interpretations, and the ability to corroborate the findings of each individual method with the other (Sandelowski, 2003). DeKeyser (2014) calls for study abroad to be investigated from a MMR perspective because as an approach it enables both reliable measurement of outcome variables and in-depth examination of learner behaviour and interaction.

MMR is not without criticism. It presents a challenge to the researcher, particularly to the sole researcher, because he/she must be competent in the use of multiple methods. Morse (2003) states that because the less dominant strategies in MMR do not constitute a complete study in themselves (e.g. in multi-method research), MMR designs are less rigorous than multi- or single-method designs. However, she also propounds that the comprehensiveness of MMR is a strength because the domain of inquiry is not limited by the method or the bias of the researcher towards one paradigm or the other.

Because it combines both QUAN and QUAL research strategies, a discussion of the epistemology of MMR would be incomplete without consideration of the theories of knowledge underpinning these two paradigms. QUAN research is traditionally aligned with a positivist perspective, whereby, as in the natural sciences, there is believed to be an objective reality, the researcher's account of what they are researching constitutes the notion of truth, and the researcher is detached from the researched. QUAL research is associated with constructivism, that is, the belief that reality is emergent, constructed by

the social actors involved, and therefore that the researcher's account constitutes their own construction (Bryman, 2004).

The present author takes a pragmatic standpoint. Pragmatism emerged in the late nineteenth-century with the work of philosophers e.g. Charles Sanders Pierce and William James. Whilst a number of different forms of pragmatism exist (Cresswell, 2003), its constant underlying principles are a focus on the practical applications and consequences of concepts. Morgan (2007) argues for an 'integrated methodology' in the social sciences based on his interpretation of pragmatism: Rather than a deductive (positivist) or inductive (constructivist) approach, he posits abduction, that is, moving between the two to observe, theorising from those observations, and acting to test the theories. He suggests an 'intersubjective' view of reality; that an objective reality exists, but that we all interpret it subjectively. Thirdly, Morgan (2007) argues for transferability as opposed to gross generalisation (QUAN) or avoidance of inference based on the argument that all phenomena are too context-specific to be generalised (QUAL). Transferability therefore is the 'appropriate use' of findings from one setting with one particular methodology, in other circumstances. Pragmatism is well suited to a study that combines measures of concrete variables (vocabulary gain and strategic behaviour) with investigation of subjective experience (context of learning) and the use of introspective research methods.

1.4.2 THEORETICAL FRAMEWORK

The primary theory of learning underpinning this study is that of Cognitivism. Cognitivist theory views learning as systematic processing of information that results in meaningful storage systems and personalised mental knowledge representation. Learning strategies play a key role in cognitivist learning theory because they represent the tools a learner uses to effectively manage and store the information they receive. Thus, as language learning

strategies are under investigation in this piece of research, many principles of cognitivism are inherent in it.

Block (2003) argued that cognitivism casts the brain as being wholly dependent upon external cues; a mechanism that is not creative beyond the stimuli to which it is exposed. In response to the mechanistic role allocated the mind by cognitivism, Situated Cognition Theory encompasses the interaction between mind and context: situated cognition theory posits that, rather than being two separate entities with the mind responding to context, mind and context are one, a whole, and cognition takes place as a result of the interaction between the two parts of the whole (Bredo, 1994). Situated cognition provides a theoretical lens through which to consider the role of informal L2 contact in this study, and to consider the strategic behaviour manifested in that contact as the product of both mind and context.

1.4.3 SUMMARY OF FRAMEWORKS OF REFERENCE

A pragmatic conceptual framework was employed to conduct this study in order to give equal prominence to generalizable and context-specific findings and to allow for transferability between data derived from QUAN and QUAL methods. Cognitivism provided the primary theoretical lens; however, in order to encompass the crucial role of context in this piece of research, situated cognition theory contributed heavily to the theoretical framework adopted.

CHAPTER II

REVIEW OF THE LITERATURE

In this chapter is detailed a critical discussion and analysis of the theories, data and research which have lain the foundations for the present study. Firstly, an analytical scrutiny of the literature on study abroad and language learning is provided. Thereafter follows an in-depth examination of the current theory and evidence with regard to vocabulary knowledge, acquisition and assessment. Next, a review of the knowledge pertaining to language learning strategies is detailed, with a specific focus on vocabulary-related strategic behaviour and the research methods used to investigate strategic behaviour to date. Finally, a synthesis of the current evidence pertaining to the three principal variables (informal second language contact, vocabulary-related strategic behaviour, and vocabulary gain) is provided, into which the present study is situated.

2.1 STUDY ABROAD EXPERIENCE

Study abroad experience (SAE) 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 (such as the ERASMUS programme in Europe). Academic SAE can last from one semester to one or two academic years. Further variation exists in non-academic SAE for language acquisition: programmes include work internships, summer language courses, intensive language courses, 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 SAEs range from a long weekend to over a year spent living and studying in the target language community.

The number of international higher education students worldwide has quadrupled in the past three decades and the UK is currently the second most popular academic SAE destination globally (after the U.S.A.), playing host to 488,380 international students in 2012 (Institute of International Education, 2013). The UK is currently the number one destination for study abroad students of English as a foreign language (EFL), benefitting from an annual influx of approximately 700,000 language learners (British Council, 2014), who earned the industry revenue of almost £2.5 million in 2012 (Norris, 2012). Popular destinations for learners of languages other than English include China, France, Canada, Japan and Spain.

Research into study abroad for language learning began with the work of John B. Carroll in 1967 in a study published in the first issue of *Foreign Language Annals*. The author had sampled n=2,784 U.S. undergraduate students in their final year of university study whose majors were foreign languages. He tested the participants on the four skills, administered the *Modern Language Aptitude Test* (Carroll & Sapon, 1958) and administered a questionnaire that asked respondents about their language learning history and their motivations, interests and attitudes toward language learning. Additionally, each of the 203 universities from which the participants were sampled completed a questionnaire designed to collect data about their approach to foreign language teaching and learning. Carroll (1967) found a very strong relationship between study abroad experience and foreign language proficiency amongst his sample. He stated, "Time spent abroad is clearly one of the most potent variables we have found...Certainly our results provide a strong justification for a 'year abroad'...a tour abroad, or a summer school course abroad" (1967:137).

Carroll's (1967) statement seeped into the 'collective knowledge' to the extent that, even today, the widely held assumption is that language learners on a SAE will have more frequent and more meaningful interaction with the L2 and that this interaction will result in a significant and positive impact on their linguistic and intercultural competence. Despite this, research into the impact of SAE on language acquisition did not begin apace until the 1990s, and particularly until after Freed's (1995) edited volume on study abroad and second language acquisition. Thereafter, scholars have sought to determine the impact of study abroad on components of language acquisition such as vocabulary acquisition (e.g. Milton & Meara, 1995; Dewey, 2008), grammatical competence (e.g. Duperron, 2006); Isabelli, 2004, 2007), oral fluency (e.g. Segalowitz & Freed, 2004, Llanes & Muñoz, 2009), writing (e.g. Sasaki 2004, 2007, 2009; Godfrey, Treacy & Tarone, 2014), reading (e.g. Dewey, 2008), listening (e.g. Dyson, 1988; Cubillos, Chieffo & Fan, 2008), strategic behaviour (e.g. Miller & Ginsberg, 1995; Carson & Longhini, 2002), narrative competence (e.g. Collentine, 2004), intercultural competence (e.g. Watson, Siska & Wolfel, 2013; Reynolds-Case, 2013), pragmatic competence (e.g. Cohen & Shively, 2007; Hassall, 2013), language learner beliefs (e.g. Tanaka & Ellis, 2003; Kaypak & Ortaçtepe, 2014) and perceptions of SAE (e.g. Wilkinson, 1998, 2002; Kinginger, 2008). Two methodological approaches pervade this research into the impact of SAE on language acquisition: (1) studies that employ a prospective, within-groups design to compare learners' pre- and post-SAE linguistic proficiency; and (2) studies that employ a between-groups design to compare SAE learners with learners in other language learning contexts in terms of their proficiency.

Another avenue of investigation in SAE research has focused on the variables that may influence linguistic gain on a SAE, which has led to a focus on individual and group differences that can mediate the impact of SAE, and the differential features of SAE

settings. In terms of individual and group differences, the factors studied have included age (e.g. Llanes & Muñoz, 2013; Dewey, Bown, Baker, Martinsen, Gold & Eggett, 2014), gender (e.g. Brecht & Robinson, 1995; Díaz-Campos, 2004), personality (e.g. Harrison & Voelker, 2008; Ozańska-Ponikwia & Dewaele, 2012), and proficiency level (e.g. Milton & Meara, 1995; Lapkin, Hart & Swain, 1995). Regarding features of the SAE context, the factors studied have included living arrangements during the SAE (e.g. Magnan & Back, 2007; Vande Berg, Connor-Linton & Paige, 2009), location of SAE programme (e.g. Dewey *et al*., 2014), length of stay (e.g. Davidson, 2010; Baró & Serrano, 2011) and type/amount of L2 contact (e.g. Martinsen, 2011; Briggs, 2014).

To sumamrise, research into SAE can be divided into two strands: studies on the impact of SAE on components of L2 proficiency, and studies on variability in the impact of SAE on L2 proficiency. Each of these two areas of investigation is hereby discussed in detail and the findings of the studies they comprise analysed and synthesised in order to arrive at a summary of the present evidence on the relationship between SAE and L2 acquisition.

2.1.1 THE IMPACT OF SAE ON SECOND LANGUAGE ACQUISITION

The research on the impact of SAE on L2 acquisition is discussed in turn with regard to the different components of language acquisition that have been researched. The impact of SAE on vocabulary acquisition and strategic behaviour is mentioned in brief in this section and approached in depth in sections 2.4 and 2.5, after the literature on vocabulary knowledge and strategic behaviour respectively has been reviewed.

2.1.1.1 Speaking

Of the research into the impact of SAE on language acquisition, the majority of studies have taken oral proficiency as their dependent variable. The findings of these studies have

for the most part suggested a benefit to oral fluency of SAE, and in particular where the gains of SAE learners have been compared with their counterparts in the home country context. The findings with regard to complexity and accuracy of oral production have been less unanimously positive.

Lennon (1990) investigated the oral fluency of four SAE learners on a six-month long sojourn in the UK to find gains in three areas of oral fluency: proportion of filled pauses per T-unit, proportion of T-units followed by a pause, and speech rate. However, his study did not include a control group and as such the role of the SAE context in these gains is unclear. Freed (1995) studied the oral fluency development of L2 learners of French in SAE and at home (AH) settings. Native speaker judges were employed to rate the spoken production of both groups from 'not very fluent' to 'extremely fluent', to find that the SAE group were deemed to produce more utterances and at a faster speed than their AH counterparts.

Yager (1998) studied the oral development of n=30 mixed-L1 learners of Spanish as a L2 on a seven-week summer abroad programme in Mexico, and n=4 control participants who remained in the U.S.A. Pre and post the SAE programme all participants completed a picture description task. Thirty-two native Spanish speakers were employed to rate the sample's pre and post-test oral samples in term of 'nativeness' on a seven-point scale from 'nada nativo' to 'completamente nativo'. A rating was decided upon for each oral sample overall, for the use of grammar in the oral sample, and also for the pronunciation. N=22 of the experimental group improved their oral proficiency in at least one of the three areas rated ($p<.05$) and n=19 made a significant improvement in all three areas ($p<.05$). The control group made no significant gains.

Segalowitz and Freed (2004) studied a group of n=47 L1 English L2 Spanish learners comprised of n=18 AH learners in the U.S. and n=22 learners in Spain over a period of one academic term. The ACTFL Oral Proficiency Interview (OPI) was administered pre and post the study period to find that the SAE group made greater gains in oral proficiency than did the AH group. Juan-Garau and Pérez-Vidal (2007) conducted a prospective study of the oral development of L1 Spanish/Catalan undergraduate students of L2 English. The researchers collected data at four time points: at the start of the university course, six months later (just prior to leaving for a SAE), on return from the SAE, and one year after the return from the SAE. The findings suggested that although SAE had an impact on the oral fluency of the sample, it did not improve their accuracy or complexity.

Llanes and Muñoz (2009) studied the impact of a short SAE on oral fluency and accuracy. They sampled n=21 L1 Catalan/Spanish learners of L2 English who studied in the target language community for 3-4 weeks. The pre and post SAE speaking test was a picture narration task. The researchers defined oral fluency as comprising six components (2009:356): (1) syllables per minute; (2) other language word ratio; (3) filled pauses per minute; (4) silent pauses per minute; (5) articulation rate (i.e. number of word spoken per minute minus the length of silent pauses); and (6) longest fluent run. Oral accuracy was operationalised as comprising the ratio of error-free clauses, average of errors per clause, and the ratio of different types of errors made (morphological, syntactic, lexical or 'covered'). However, the definition given by the researchers for 'covered' errors does not clearly distinguish this type of error from lexical errors. In terms of fluency, Llanes and Muñoz's (2009) participants made gains on syllables per minute [t=6.108, df=23, p=.000], other language word ratio [t=4.383, df=23, p=.000], articulation rate [t=4.556, df=23, p=.000] and longest fluent run [t=2.493, df=23, p=.020]. With regard to accuracy, the sample made significant gains in terms of ratio of error-free clauses [t=2.168, df=23,

p=.041], average of errors per clause [*t*=3.314, df=23, *p*=.003], and lexical errors [*t*=4.513, df=23, *p*=.000]. However, due to the relatively small sample size, a 90% confidence interval was found. Moreover, without corresponding data from a control group, conclusions cannot be drawn as to the role of SAE in the gains reported.

Not all studies of the impact of SAE on oral proficiency have found in its favour. For example, Freed, Segalowitz and Dewey (2004) compared three groups of L2 French learners (SAE, EFL and immersion) regarding oral fluency to find that the SAE group made fewer gains over one semester than the immersion group. Vande Berg, Connor-Linton and Paige (2009) investigated the impact of living arrangements on the oral proficiency gains of n=830 American learners of a range of L2s across 61 study abroad programmes. The researchers tested for oral proficiency pre and post the SAE using the Simulated Oral Proficiency Interview (SOPI), which is very similar to the ACTFL OPI. They found no significant difference between the gains of participants who lived with other U.S. natives and participants who lived with L2 native speakers.

2.1.1.2 Writing

The empirical evidence on the relationship between SAE and writing development is inconclusive to say the least. Freed, So and Lazar (2003) studied the writing fluency development of n=30 learners of L2 French, half of whom studied abroad and half who did not. They administered a writing test (an essay task) pre and post the SAE in France. Unfortunately, the researchers only analysed the writing of n=8 participants: n=4 from the SAE group and n=4 from the AH group. Native-speaker judges rated the fluency of the compositions on a seven-point scale from 'not at all fluent' to 'very fluent'. The researchers gave no definition of the term 'writing fluency' to the raters; indeed, the researchers state that their directions to the raters gave them free rein to interpret the term

'fluency' as they saw fit. As such, the rating system (and consequently, the study) is invalidated because the raters may each have interpreted the term very differently and thus given their ratings in response to different aspects of the compositions. Additionally, no mention is made of whether the essay tasks were timed; a bizarre omission given that writing fluency is often operationalised as rate of composition (e.g. Hatasa & Soeda, 2000; Chenoweth & Hayes, 2001). The raters were provided a checklist of textual features (e.g. organization, vocabulary, complexity) which they were required to tick to indicate which aspects of each composition had most strongly influenced their rating of it. Moreover, Freed, So and Lazar (2003) analysed the texts in terms of complexity and lexical density by calculating the length of T-units. They found no significant difference between the gains of the groups from time 1 to time 2 as rated by the native speakers, nor in gains on mean length of T-unit as measured by the researchers. The only difference between the groups was that the SAE learners made greater gains in length of composition. This finding suggests that, had the researchers operationalised the construct of fluency as number of words per minute, they may have found an impact of SAE in comparison with the at home setting.

Pérez-Vidal and Juan-Garau (2009) sampled n=37 L1 Spanish, advanced proficiency level learners of L2 English in a prospective study of the development of written proficiency as a result of SAE. The researchers tested their sample at four time points: at the beginning of the study, six months later (during which time the participants had received 80 hours of formal English instruction), six months after time 2 (at which point the sample had just returned from a six-month SAE in the UK that had not included explicit instruction in L2 writing), and finally fifteen months after the return home from the SAE. At each time point the sample were required to produce a written composition and to complete a self-report questionnaire. The compositions were analysed for fluency (operationalised as words per

clause and words per minute), accuracy (ratio of errors per word), and complexity (number of dependent clauses per dominant clause and 'coordination index', which refers to a division of the total independent clause coordinations by the total combined clauses). Moreover, Guiraud's (1954) *Index of Lexical Richness* (a ratio of types and tokens in a composition) was used to measure lexical complexity. They found that when fluency was measured by words per clause, there was no significant difference between the compositions at time 2 and time 3 (i.e. across the SAE period). However, when measured by words per minute there was an effect of time on fluency [$F(1, 36)=27.23$, $p<.000$]. There was no significant effect of time spent abroad on accuracy or complexity except on Guiraud's index: the sample increased their lexical richness during the SAE [$F(1, 36)=17.54$, $p<.000$]. This finding mirrors those found by Sasaki (2004, 2007, 2009), who studied the writing development of SAE and AH learners over three and a half years to conclude that the differences in writing competence between the groups was due to a larger vocabulary size on the part of the SAE learners.

Godfrey, Treacy and Tarone (2014) conducted an investigation into the L2 writing development of n=8 third-year American undergraduate students of French over one academic term. N=4 of the participants spent the term at the University of Minnesota; n=4 spent the term at the Université de Montpellier III–Paul Valéry III. Pre and post the academic term all participants took the ACTFL computer-based OPI and the ACTFL Writing Proficiency Test. Additionally, at both time points the sample completed a timed writing task in response to a photo prompt. The writing tests were analysed for complexity, accuracy, fluency, and for form-function proficiency with reference to making and supporting claims. The data revealed mixed findings: all of the SAE participants moved up a level between the pre and post ACTFL tests in both writing and speaking, whereas only one of the AH participants moved up a level, and only on the writing test. The results

pertaining to fluency (measured by length of composition within the 50-minute time limit) and complexity (measured in T-units) suggested that the AH group made greater gains than the SAE group. The SAE group made greater gains on accuracy (measured in terms of gendered nouns and gender marking), and the AH group made greater gains in form-function proficiency (measured in terms of proportion of hedges when making claims).

Why might Godfrey, Treacy and Tarone's (2014) AH group have made more gains in writing proficiency than their counterparts in France? It may be the case that the SAE participants received less input on writing during the academic term than did the AH group: without details about the SAE/AH programme features or the language contact experienced by the sample over the course of the term, only suppositions can be made. The findings of this study may not be externally valid due to the small sample size and the resulting dearth of statistical power. Moreover, their measures of writing proficiency are not comprehensive; for example, many scholars have argued that the T-unit is insufficient to measure all aspects of complexity (e.g. Crookes, 1990; Biber, Gray & Poonpon, 2011).

2.1.1.3 Listening

Some of the studies of the impact of SAE that have taken listening comprehension as the dependent variable have found evidence of significant gains. For example, Llanes and Muñoz (2009), in their study of L1 Catalan/Spanish learners of L2 English on a short SAE, administered a listening comprehension test before and after the stay abroad. In the listening test the participants were required to choose one picture from a choice of three that best illustrated each of the 21 stimulus sentences that they heard. A *t*-test revealed that the gains the sample made from time 1 to time 2 were statistically significant [t=5.541, df=23, p=.000]. Two other studies that found a significant impact of SAE on listening both sampled learners of French: Allen and Herron's (2003) n=25 participants spent six weeks

in France and made significant gains pre and post the sojourn on a multiple-choice listening test. Similarly, Kinginger's (2008) n=23 participants made significant gains on the *Test de Françias International* after an academic term in the target language community. Unfortunately, none of these three studies included a control or comparison group so, whilst the findings are suggestive of an impact of SAE, they are by no means conclusive.

Other studies of SAE and listening have failed to uncover any relationship between them. For example, Huebner (1995) conducted a study into the impact of a nine-week summer SAE in Japan on the listening proficiency of n=12 American, beginner-level learners of Japanese. He administered the *Educational Testing Services' Japanese Proficiency Test* pre and post the SAE, and also to n=12 learners of Japanese who spent the summer studying Japanese in a domestic (at home) immersion setting. No significant differences were found between the SAE and at home groups in terms of their listening skills. Similarly, Tanaka and Ellis (2003) sampled n=166 L1 Japanese learners of English, who had enrolled in a 15-week long SAE in the U.S., to investigate the impact of the stay abroad on listening comprehension skills. The researchers administered the listening section of the TOEFL pre and post the SAE to find minimal gains. Moreover, in their study of the listening comprehension gains of n=48 L1 English learners of L2 Spanish, Cubillos, Chieffo and Fan (2008) found that only those SAE participants whose initial proficiency level was high made significant gains in comparison with their AH counterparts.

2.1.1.4 Reading

Of the four skills, reading has received the least amount of attention from the study abroad research community. What empirical evidence exists is as yet too tentative to suggest a

relationship between a period abroad and increased reading proficiency. Kline (1998) conducted an ethnographic study of the reading development of a learner of French in France to find that the host family, social network and gender influenced the types of L2 texts they chose to read during the SAE. Huebner (1995) studied n=12 American, beginner-level learners of Japanese on a nine-week SAE in Japan. He administered the Educational Testing Services' Japanese Proficiency Test pre and post the SAE, and also to n=12 learners of Japanese who spent the summer studying Japanese in a domestic (at home) immersion setting. The SAE learners made greater gains than the domestic immersion group but not in all reading measures: there was no significant difference between them on character recognition. Moreover, the SAE group expressed a strong desire to improve their reading because they felt they were inhibited by their lack of skill, whereas the at home learners stated that they were frustrated by reading in Japanese and felt that having to learn two syllabaries was an obstacle to their success.

In terms of studies that did not find in favour of SAE for reading, Díaz-Campos (2004) found no significant difference between the reading performance of a SAE and an AH group after an academic term abroad. In a study with similarly disappointing findings, Dewey (2004) investigated the reading comprehension development of n=30 L1 English, L2 Japanese learners, half of whom were on a SAE in Japan and half on a domestic immersion programme in the U.S.A. He used think-aloud protocols pre and post the SAE period to measure free recall (the percentage of clauses a participant was able to recall from within one passage) and also administered a vocabulary test (whereby respondents provided definitions for 30 words randomly selected from the reading passages used during think-aloud) and a self-assessment questionnaire. Dewey (2004) found that the SAE group outperformed the AH group only on their self-assessment of reading comprehension proficiency: the SAE learners were more confident about their reading than their AH

counterparts. There were no significant differences between the groups on free recall or vocabulary knowledge. Huebner (1995) and Dewey (2004) both found that there was a higher level of variation in the post-test reading scores of the SAE participants in comparison with the AH and domestic immersion groups respectively.

2.1.1.5 Grammar

There is little empirical evidence of the impact of SAE on grammatical competence and the findings of the studies that have been conducted are contradictory. Some research is in agreement that there is no significant impact: DeKeyser (1991) and Collentine (2004) both found no significant difference post SAE for grammatical competence. Collentine (2004) studied n=46 L1 English learners of L2 Spanish, of whom n=26 spent an academic term on a SAE in Spain. He analysed pre and post SAE OPI data in terms of 17 morphological, syntactic and morpho-syntactic subcomponents of grammatical control, and compared the SAE group data with that of their AH counterparts. Collentine (2004) found that the AH participants outperformed the SAE group on the five subcomponents that significantly distinguished between them. DeKeyser (1991) also compared the grammatical gains of a group of n=12 American learners of L2 Spanish to find that the SAE group were indistinguishable from the AH group after a six-month sojourn in Spain. However, the small sample size in his study suggests that this finding is not generalizable. Isabelli-García (2010) studied gains with regard to gender agreement of n=12 SAE learners of L2 Spanish over one academic term in comparison with n=12 at home learners. She administered a grammaticality judgment test pre and post the academic term to find no significant difference between the groups. The small number of participants in each group may have limited the power of the statistical analyses to discriminate between them, and the researcher suggests that the findings may be attributable to a greater focus on form in at home settings than on a SAE.

Other researchers have found evidence to suggest an impact of SAE on grammatical competence, yet the SAE of their samples has tended to be longer than the SAEs of DeKeyser (1991) and Collentine's (2004) samples, suggesting that length of stay is a mediating factor. For example, Guntermann (1995) studied the grammatical development of n=9 L2 learners of Spanish, whose SAE was a year in length in Central America. Oral interviews were analysed to reveal an improvement in the use of the copular verbs *ser* and *estar* and in use of the past tense. Similarly, Isabelli (2004) studied n=29 L2 learners of Spanish who spent a year in Spain to find significant gains at the post tests (a grammaticality judgment test and an oral interview) in terms of syntactic development in the areas of subject pronoun omission and subject-verb inversion.

2.1.1.6 Vocabulary

All of the studies conducted to date on lexical development as a result of SAE have found evidence of gains. Studies of lexical development have taken a prospective, within-subjects design (e.g. Milton & Meara, 1995; Ife, Vives Boix & Meara, 2000; Llanes & Muñoz, 2009) or a between-groups methodology that compares the gains of a SAE group with participants in other settings (e.g. Collentine, 2004; Dewey, 2008; Foster, 2009; Serrano, Llanes & Tragant, 2011). The studies that have specifically addressed vocabulary gain in a study abroad context are attended to in detail in section 2.4.1 and the weight of their evidence evaluated in light of the research aims of the present study.

2.1.1.7 Pronunciation

Research into the impact of SAE on phonological control is somewhat scarce, and its findings so far inconclusive. Yager's (1998) study of n=30 learners of L2 Spanish on a sojourn in Mexico found significant gains in native-like pronunciation for the majority of the SAE participants but none for the control group who studied in their home countries.

Similarly, Díaz-Campos (2004) found that SAE learners made significantly greater pronunciation gains than their AH counterparts. Moreover, Avello, Mora and Pérez-Vidal (2012) found that a SAE led to a decrease in the proportion of their sample's pronunciation errors.

In another study that found in support of an impact of SAE on pronunciation, Muñoz and Llanes (2014) investigated the impact of SAE on the degree of foreign accent. They sampled n=55 L1 Spanish learners of L2 English, of whom n=28 were children under 12 years of age and n=27 were adults (over 18 years of age). Moreover, the sample was divided into learners of English as a foreign language at home in Spain (the AH group) and learners of English in the UK (the SAE group). The AH group comprised n=13 children and n=15 adults and the SAE group comprised n=15 children and n=12 adults. Pre and post-SAE tests were administered one week prior to and one week following the SAE, and the same tests were administered to the AH group with the same interval of time in between. Testing comprised a semi-structured interview for demographic and background data, a picture narrative task and, at time 2 only, a questionnaire that collected data on L2 contact between the pre and post-tests. This questionnaire is remarkably similar to the Language Contact Profile (a questionnaire designed by Freed, Dewey, Segalowitz & Halter, 2004; c.f. 2.1.2.3a) in that it requires respondents to state how many days per week and hours per day they engaged in different types of contact with the L2. 28 raters listened to excerpts from the narration tasks and rated the degree of foreign accent on a 7-point Likert scale from 'no foreign accent' to 'very strong foreign accent'. Inter-rater reliability was very high at >.98. A 2 x 2 ANOVA that took age (child; adult) and learning context (AH; SAE) as the independent variables found no effect for age but a significant impact of context: the SAE participants made greater gains in terms of lessening the degree of their foreign accent than did the AH group [$F(1, 51)=6.852$, $p=.012$, $\eta^2=.118$]. The researchers

attribute this finding to the greater amount of time spent speaking with native speakers reported by the SAE group.

Not all studies have found in favour of the impact of SAE on phonological control. For example, over a period of three years Mora (2014) measured the perceptual phonological gains of n=66 bilingual Spanish/Catalan learners of L2 English. Data was collected at four time points but participant mortality meant that only n=27 completed the testing measures at time 4. The four time points were: (1) at the beginning of the academic year; (2) three months later at the end of the first term, during which three months of formal English instruction had been delivered (and just before the start of the SAE); (3) three months after that (at the end of the SAE); and (4) fifteen months after the end of the SAE. The phonological perception test was an auditory discrimination task that required respondents to differentiate between word-final consonant pairs (voiced or unvoiced) and between vowel pairs. The findings suggested that significant gains only occurred between time 1 and 2 (the period in which at home formal instruction was delivered). As such, the data suggested that at home instruction has an impact on perceptual phonological competence, whereas SAE does not.

2.1.1.8 Strategic behaviour

There is little evidence on whether and how SAE impacts upon language learning strategies, but what qualitative data does exist suggests that SAE has no significant impact on the strategies a learner employs (Huebner, 1995; Miller & Ginsberg, 1995). The evidence to date on strategic behaviour in a study abroad context is discussed in detail in section 2.4.2, after the literature on strategic behaviour has been reviewed.

2.1.1.9 Narrative competence

One study has examined the relationship between SAE and narrative competence to find a positive relationship between the two. Collentine (2004), in his study of n=46 L1 English learners of L2 Spanish in SAE and AH settings, administered an OPI to his sample before and after a semester-long SAE. Taking a discourse analysis approach to the interview transcripts, he analysed the data in terms of the features of narrative discourse as put forth by Biber (1988): 'public' verbs (that is, verbs which refer to events and/or communication), past and present participles, past tense verbs, and third-person morphology. Collentine (2004) counted the number of instances of each of these features in each transcript and added them to arrive at a narrative score to be used in repeated measures comparisons. However, he does not state if he counted only accurate usage of the features or included both accurate and erroneous uses. ANCOVA revealed that the SAE group used significantly more narrative features at time 2 than their AH counterparts, and that the gain between their time 1 and time 2 narrative scores was significant [$F(1, 43)=5.00$, $p=.031$, $r^2=.445$].

2.1.1.10 Pragmatic competence

Studies that have focused on the relationship between SAE and pragmatic competence have tended to operationalize the dependent variable in very disparate ways from one another. In a qualitative study of SAE and pragmatic competence, Shively (2013) investigated the development of L2 Spanish conversational humour in n=6 U.S. study abroad learners. Over the course of one academic term she recorded eight interactional conversations between each participant and either their Spanish host family or a L2 native-speaker friend. Shively (2013) found that her sample did engage in humorous L2 talk and that they utilized the linguistic and cultural knowledge they had acquired during the SAE

as aids to this end. Moreover, she found that the participants became increasingly aware of the difference in the use of sarcasm between the L1 and L2 to the extent that they either avoided sarcasm in the L2 altogether or modified their use of it in order to garner more favorable responses to their attempts at humour.

In a vastly different approach to Shively (2013), Taguchi (2008) sampled n=44 L1 Japanese SAE learners of L2 English in an investigation of the development of their pragmatic competence over the four months of their stay in the target language community. The construct was operationalised as the speed and accuracy with which the sample was able to recognise indirect refusals and opinions in a listening task. The findings revealed that only the speed of comprehension of these pragmatic features improved; accuracy of comprehension was not improved.

In another very different operationalisation of pragmatic competence, Hassall (2013) examined the development of the use of address terms of n=12 Australian L2 learners of Bahasa Indonesia on a short SAE in Central Java. He administered a pre and post test of perceptions about L2 address terms, in which the respondent is given an imaginary situation and is required to choose which address term they would use in that scenario. Moreover, the sample kept learner journals during their stay, and they were interviewed at the end of the SAE. The study found that over the course of the seven-week SAE the sample rapidly gained a number of address terms for use in the vocative slot (i.e. 'Good afternoon, *Sister*') but did not make as much progress with regard to terms for use in the pronoun slot and seemed less concerned about conforming to native speaker norms in this regard. Similarly to Hassall (2013), Reynolds-Case (2013) employed a measure that required respondents to formulate an L2 address form in response to a hypothetical scenario; in this case, the Spanish pronouns *ustedes* and *vosotros*. Her n=10 L1 English sample spent four weeks in Madrid and in that time she observed them daily. Her results

suggested that the sojourn abroad had sensitised the participants to the situations in which the more formal *ustedes* is appropriate.

2.1.1.11 Intercultural competence

As with pragmatic competence, studies that have explored SAE in terms of intercultural competence have come to disparate conclusions, most likely due to their differing conceptualisation of intercultural competence. Paige, Cohen and Shively (2004), in a study of SAE learners in France and Spain, found evidence to support a positive impact of SAE on cross-cultural competence, and Reynolds-Case's (2013) study of n=10 L1 English learners of L2 Spanish who spent four weeks in Madrid found that the sample were able after their SAE to distinguish between more and less formal scenarios in the L2. Conversely, Medina-López-Portillo (2004) found no significant gains in intercultural competence for learners of Spanish after a seven-week SAE in Mexico. Moreover, Isabelli-García's (2006) study of SAE learners of L2 Spanish in Argentina found that learners who harboured negative feelings about the target language culture were less motivated to interact with native speakers. Additionally, it was not always the case that positive feelings toward Argentine culture developed as a result of studying there.

Watson, Siska and Wolfel (2013) studied the intercultural competence development of n=498 undergraduate students at a US military academy who spent an academic term in the target language community of a range of L2s (Arabic, Chinese, French, German, Portuguese, Spanish or Russian). The four pre and post SAE measures comprised: (1) a test of regional knowledge, which tested knowledge of the physical, human and cultural geography of each region visited; (2) a regional aptitude network test, which measured region-specific sensitivity to cultural scenarios and interrelationships; (3) the *Intercultural Development Inventory*, which is a self-report measure of intercultural competence; and

(4) a language proficiency test. The findings revealed that the sample made significant gains in cross-cultural competence (yet the researchers did not report the significance value of this finding), but only the participants who studied abroad in China or Eastern Europe made significant gains in regional knowledge. The scope of this study was so wide in terms of L2s, SAE programmes and destinations, and so global in terms of its measures, that although the findings appear to support the impact of SAE on intercultural competence, a vast number of variables (e.g. amount and type of L2 contact, living arrangements, location) may have potentially confounded the results.

2.1.1.12 Language learner beliefs

Empirical evidence suggests that a stay in the target language community changes language learning beliefs, and particularly beliefs that relate to the self as a language learner. For example, a survey conducted by the Institute for the International Education of Students and reported by Dwyer (2004), which yielded almost 3,500 responses, indicated that after returning home from a SAE, learners' self-confidence as a L2 user was higher than before their stay abroad and was sustained beyond the end of the SAE. In a study which yielded similar findings, Tanaka and Ellis (2003) studied the beliefs about language learning of n=166 L1 Japanese learners of L2 English who were on a fifteen-week sojourn in the U.S. The researchers administered a 27-item questionnaire on language learning beliefs, which required respondents to rate on a five-point scale how strongly they agreed or disagreed with the 27 statements. The questionnaire was administered three weeks before and three days after the SAE period. The data was analysed with reference to three sub-components of language learner beliefs: (1) beliefs about the self; (2) beliefs about learning analytically; and (3) beliefs about learning experientially. All three categories of beliefs were significantly and positively impacted by the stay in the U.S. (p<.001), with the strongest gains in beliefs about self, particularly beliefs about self-efficacy and

confidence. The researchers also administered the TOEFL pre and post the SAE but found no correlation between the gains in beliefs and gains in L2 proficiency. Although this study points to a clear impact of SAE on beliefs, it must be noted that the sample lived together whilst in the U.S. and also studied together in monolingual classes with a Japanese teacher, so how far their experience of SAE was different from their experience of AH L2 instruction is unclear.

Amuzie and Winke (2009) sampled n=77 students on a SAE in the U.S who came from a variety of countries and represented a range of L1s (Korean, Vietnamese, Chinese, Arabic, Japanese and French). The sample were divided into two groups based on length of stay: Group 1 had been in the U.S. for less than six months at the point of data collection (M=4.11, SD=1.83), and Group 2 had been in the U.S. more than six months (M=14.19, SD=5.87). The researchers administered a language learning beliefs questionnaire that utilised a 10-point response scale. Qualitative data was collected via interviews in which each participant's questionnaire responses were used as stimuli. Factor analysis of the questionnaire data revealed three underlying factors: beliefs about the role of the teacher, beliefs about learner autonomy, and beliefs about self-efficacy. One significant difference was found between the groups: in a finding which mirrored that of Tanaka and Ellis (2003), the longer stay group had stronger beliefs about self-efficacy [$t(68)=2.39$, $p=.02$].

2.1.1.13 Perceptions of SAE

There is evidence to suggest that language learners' perceptions of SAE are sometimes negatively impacted by the SAE itself. For example, in an ethnographic approach to researching study abroad, Wilkinson (2002) used conversation analysis to examine the interactions between n=4 North American learners of French in France and the host families they were living with. One of her participants reported feeling patronised by her

host family and by extension she avoided the company of L2 native speakers and sought out fellow Americans to converse with instead. Kinginger (2008) found that the less demanding nature of the work on an academic SAE as compared with the AH academic setting led to one of her participants viewing the study abroad setting negatively as less academically rigorous. Another of Kinginger's (2008) participants experienced a change in perception as a result of having studied abroad: this participant took the SAE and the challenges it had presented her with in terms of integrating into French society as proof that American cultural and linguistic norms were superior to the those of the target language community.

2.1.1.14 Summary of the impact of SAE on SLA

The literature here discussed on the impact of SAE on second language acquisition has suggested that SAE learners are significantly more orally fluent after their SAE in comparison with learners who studied in the home country context (DeKeyser, 1991; Freed, 1995; Lafford, 2004; Segalowitz & Freed, 2004). Moreover, SAE learners experience a significant impact of SAE on the development of their vocabulary knowledge (Milton & Meara, 1995; Lennon, 1990; DeKeyser, 1991; Ife *et al*., 2000; Dewey, 2008; Llanes & Muñoz, 2009). However, for other components of second language competence the benefit of SAE over EFL settings has not yet been firmly established (Lennon, 1990; Collentine, 2004; DeKeyser, 1991; Freed *et al*., 2003; Díaz-Campos, 2004; Dewey, 2004; Mora, 2008).

2.1.2 VARIABLITY IN THE IMPACT OF SAE

A great many studies (e.g. Brecht, Davidson & Ginsberg, 1995; Freed, 1995; Brecht & Robinson, 1995; Ife, Vives-Boix & Meara, 2000; Magnan & Back, 2007; Harrison & Voelker, 2008; Llanes and Muñoz, 2009, 2012; Dewey, Bown, Baker, Martinsen, Gold &

Eggett, 2014) have sought to determine the variables that mediate the impact of SAE on linguistic gain. The variables that have been investigated are individual and group differences such as age, gender, personality, initial proficiency level, and specific features of the SAE context such as living arrangements, location of SAE programme, and length of stay. Moreover, type and amount of second language contact has been a key issue of consideration in determining the cause of variability in the impact of SAE. These variables are now discussed in turn and the weight of evidence pertaining to each evaluated.

2.1.2.1 Individual and group differences

2.1.2.1a *Age*

In terms of immigration to a target language community, there is evidence to suggest that the earlier a language learner arrives, the better in terms of eventual mastery of the language (e.g. Johnson & Newport, 1989; Birdsong & Molis, 2001; Birdsong, 2006). As stated by Dewey *et al*. (2014), most of the studies that investigate age and study abroad have compared children with adults, but empirical evidence has been found in support of a lower age of acquisition resulting in higher levels of ultimate L2 mastery, even among adult learners (e.g. Stevens, 1999). However, the findings with regard to age and SAE are not unanimously in favour of the impact on the former of the latter. For example, Llanes and Muñoz's (2009) study of n=24 L1 Spanish/Catalan learners of L2 English on a three-to-four week stay in the target language community took age as an independent variable. The participants, whose ages ranged from 13 to 22 years old, completed a pre and post SAE oral interview, which was analysed for measures of fluency and accuracy. In multiple regression, where initial proficiency level was controlled for, there was no significant influence of age on oral fluency or accuracy outcomes.

In another study of age and SAE conducted by Llanes and Muñoz (2013), a sample of n=139 L1 Spanish learners of L2 English was divided into four groups: n=39 SAE children, n=34 AH children, n=46 SAE adults, and n=20 AH adults. Oral and written pre and post test data were collected from the sample and analysed for fluency, lexical density and accuracy. MANCOVA revealed differences between the age groups were statistically significant for the writing measures, and that on those measures the adults significantly outperformed the children. There was a significant interaction between age and learning context in terms of the writing measures [F(4, 123)=4.209, *p*=.003, Wilks's Lambda=.880, η2=.120], which revealed that the AH adults gained the most, followed by the SAE adults, the SAE children and the AH children respectively. No significant interaction effects were found for oral proficiency. The researchers suggest that the reason for this finding may lie in the less cognitively demanding nature of the oral task in comparison with the written task; that is, they believe that the further developed literacy skills of the adults may have had positive transfer on their writing proficiency scores, whereas the simple oral picture narration task favoured neither age group.

Dewey, Bown, Baker, Martinsen, Gold & Eggett (2014), in their study of n=118 L1 English SAE learners across the globe, included age as an independent variable in their sample, which comprised learners from age 18 to 26. Age was found to be a highly significant predictor of L2 contact, with older participants having more overall contact, more interactive contact and more receptive L2 contact than their younger counterparts. This finding suggests that findings such as those uncovered by Llanes and Muñoz (2013) may be as a result of the differences in language contact between learners of different ages, as opposed to stemming from the differing cognitive or literacy skills of adults in comparison with children. Clearly, more research is required in order firmly to pin down the role of age in the variability of the impact of SAE on second language acquisition.

2.1.2.1b *Gender*

With regard to gender and SAE, there is evidence to suggest that women make smaller gains than men in learning contexts in which the role of women is markedly different from their role in the home setting. Brecht, Davidson and Ginsberg (1995) conducted a longitudinal, mixed methods study into the oral proficiency gains of n=658 U.S. learners of Russian in Russia to find that gender was a very strong predictor of gain: male learners were more likely than female learners to attain an advanced proficiency level. The researchers hypothesised that the way the target language community treated the female participants may have been the root cause. That is, they suggested that the different role of women in Russian culture as compared with American culture may have impacted upon the types of interactions with the L2 that the female participants experienced, which in turn inhibited their oral L2 development. A qualitative analysis of the learner journals completed by Brecht *et al*.'s (1995) sample was conducted by Polyani (1995) to find that male learners were encouraged in L2 interaction to participate and express their opinions fully. For women, however, participating in L2 interaction was reported at times to provoke negative responses and in some cases sexual harassment from native speaker interlocutors.

In a qualitative investigation of one U.S. participant who had made large gains on oral proficiency on a SAE in Argentina and one who had not, Isabelli-García (2006) identified a number of gendered experiences that she believes may have played in a role in the smaller gains of the female learner. The learner's journal revealed that she became increasingly isolated from Argentine culture during her SAE, partly as a result of numerous experiences of being catcalled by Argentine men about her weight. Dewey *et al*. (2014) found that gender was not a strong predictor of L2 language contact in their study of n=118 L1 English learners of a range of language in a variety of host countries, yet the

SAE programmes their sample were enrolled in required their students to spend a set amount of time each day engaged in oral interaction with native L2 speakers.

2.1.2.1c *Personality*

The studies so far conducted on the impact of personality factors in the study abroad context have tended to take as the dependent variable contact with the host culture or native speakers as opposed to linguistic gain. For example, Ozańska-Ponikwia and Dewaele (2012) conducted a study of Polish learners of English who were immigrants in Ireland. Their findings suggested that Openess was the strongest predictor of L2 proficiency, and that Openess and Self-Esteem were the strongest predictors of L2 contact. Conversely, Dewey, Bown, Baker, Martinsen, Gold and Eggett (2014) failed to find an effect of personality on language contact. Harrison and Voelker (2008) set out to investigate the adjustment to the host culture of study abroad participants who had differing levels of two particular personality features: *Entrepreneurial Attitude Orientation*, which they define as the ability to recognise and/or take advantage of opportunities, and emotional intelligence. They sampled n=191 North American undergraduate students and found that emotional intelligence was a strong predictor of general cultural adjustment, but less so for adjustment to interacting with the host culture. With regard to Entrepreneurial Attitude Orientation the findings were more mixed: participants with higher levels of Entrepreneurial Attitude Orientation adjusted more to interaction, but there was no significant difference between the groups in terms of their general adjustment to the host culture.

Dewey, Ring, Gardner and Belnap (2013) studied the social network formation of n=71 SAE learners of L2 Arabic who were either studying in Jordan or studying in Egypt. The researchers took a mixed methods approach: they administered a questionnaire to

determine the extent of their participants' social networks, which also comprised free-write sections to elicit data about how these networks were managed. Moreover, the sample completed weekly learner journals about their contact with the L2 during the SAE. The qualitative data was analysed using a grounded theory approach, and personality was found to be a key theme that emerged. For example, 'personality issues' was the second most commonly cited reason in both contexts for facilitating and inhibiting deeper friendships with L2 native speakers. Although these studies point to an indirect relationship between personality and linguistic gain in a study abroad context, the evidence is so far suggestive as opposed to convincing.

2.1.2.1d *Initial proficiency level*

The literature on initial proficiency level (that is, the proficiency level of a leaner at the outset of their SAE), has tended to suggest that higher proficiency learners make less gain as a result of SAE than their less proficient counterparts. For example, Llanes and Muñoz (2009), in their study of the impact of a short SAE on listening comprehension and oral proficiency, employed regression analysis to find that for their sample of L1 Spanish learners of English, initial proficiency level was a significant predictor (p=.027) of gains in syllables per minute, ratio of words in another language, and proportion of lexical errors: the lower the initial level, the higher the gains in these features of oral competence. Similarly, Milton and Meara (1995), in their study of the vocabulary acquisition of mixed L1 participants (German, Italian, Spanish and French) on a six-month academic exchange at a British university, found that the participants who had the lowest vocabulary scores before the SAE were those who made the highest gains.

Pizziconi (2013) investigated the lexical development of n=48 L2 learners of Japanese at a UK university, whose degree course included a stay in Japan. One group of participants

went to Japan for the second year of their degree (SA2), the other for their third year (SA3). Before the year in Japan the SA2 group received intensive Japanese instruction of a 'beginner' syllabus and the SA3 group a more extensive approach which covered the beginner syllabus and part of an 'intermediate' syllabus. It is not stated explicitly whether these syllabi labels refer to the general L2 proficiency level of the participants, but from Pizziconi's (2013) interpretation of the findings we can infer that the SA2 group had a lower initial proficiency level than the SA3 group. She administered her vocabulary test immediately before the SAE for each group, immediately after for each group, and 1 year after the end of the SAE for each group. A 2 x 2 ANOVA that took group (SA2; SA3) and time of testing as the independent variables revealed that the SA2 group made greater gains than the SA3 group [$F(1, 25)=7.79$, $p=.01$]. Similarly to Milton and Meara (1995) and Llanes and Muñoz (2009) therefore, those participants with a lower initial level gained more as a result of the SAE.

The theoretical position underpinning empirical findings in favour of a lower initial proficiency level is that there may be a proficiency threshold beyond which SAE learners cease to benefit as richly from the period abroad (Lafford & Collentine, 2006). Indeed, scholars have attempted to define which particular subcomponents of linguistic competence hold the greatest sway in this threshold. For example, Golonka (2006), in a study of L1 English learners of Russian, found evidence to suggest that control of grammar, vocabulary and accuracy, and sentence repair and self-corrections, were predictive of a move from the intermediate to the advanced proficiency level on the ACTFL scale as a result of a SAE. However, not all of the findings with reference to initial proficiency level are quite as straightforward. For example, Yager's (1998) study of the oral proficiency gains of n=30 learners of Spanish on a SAE in Mexico found that for participants at intermediate proficiency level there were no significant correlations

between L2 contact and oral gains. However, for beginner and advanced level participants there were significant positive correlations between overall amount of L2 contact and gains in 'nativeness' of Spanish speaking overall [r=.79, p<.05] and of Spanish pronunciation [r=.78, p<.05]. Moreover, for both beginner and advanced level participants, significant negative correlations were found between solitary L2 contact and speaking overall, and between solitary contact and 'nativeness' of grammar usage (p<.05). Ife, Vives Boix and Meara (2000), in their longitudinal study of the impact of SAE and proficiency level on lexical development, measured the type and rate of L2 Spanish vocabulary acquired by n=36 L1 English speakers who were studying Spanish in Spain. The sample was divided into two initial proficiency level groups: intermediate and advanced. Unfortunately, the researchers fail to report how proficiency was measured in this study. However, in a finding which contradicts those of Milton and Meara (1995), Llanes and Muñoz (2009) and Pizziconi (2013), both proficiency groups made a statistically significant gain in both vocabulary measures between time 1 and time 2 (p<.013 in all cases) and there was no statistically significant difference between the gains of the two groups. Therefore, whilst there is some evidence to support a threshold hypothesis, it is by no means a given and one could argue that further research is necessary to determine its existence before more research into its possible location is undertaken.

2.1.2.2 Features of the SAE context

2.1.2.2a *Living arrangements*

Vande Berg, Connor-Linton and Paige (2009) investigated the impact of living arrangements on the oral proficiency gains of n=830 American L1 English learners of a range of L2s across 61 study abroad programmes. The researchers tested for oral proficiency pre and post the SAE using the Simulated Oral Proficiency Interview (SOPI),

which is very similar to the ACTFL OPI. They found no significant difference between the gains of participants who lived with other U.S. natives and participants who lived with L2 native speakers. However, the participants who were learning languages that use a Latin-based writing system (French, German and Spanish) gained more oral proficiency than those learning languages that use a different writing system from English (i.e. Arabic, Japanese, Chinese and Russian). For the learners of Latin-based languages, there was a statistically significant correlation between oral proficiency gains and amount of time spent with a host family, suggesting that a key variable in gain is the take up of native speaker interaction offered by a home stay living arrangement.

In a study of n=24 American undergraduate L2 learners of French on an 18-month long SAE in either Montpellier or Paris, Magnan and Back (2007) investigated the effect on oral proficiency gain of living arrangements and informal L2 contact. Before the SAE all participants completed a questionnaire about their expectations of the sojourn abroad, the Can-Do Self Assessment Scale (Clark, 1981; a self-report measure of proficiency and competence in the L2), and the ACTFL Oral Proficiency Interview (OPI). The post-SAE measures consisted of a questionnaire about their experience of the sojourn abroad, the Can-Do Self-Assessment Scale, the OPI and a version of the Language Contact Profile (LCP; Freed, Dewey, Segalowitz & Halter, 2004; c.f. 2.1.2.3a), which the authors had adapted for use with learners of French and shortened slightly in terms of numbers of items. N=11 of the sample lived with native French speakers (e.g. with a host family) and n=9 lived with non-native French speakers (e.g. in halls of residence). A Mann-Whitney U test indicated that there was no significant difference between the oral proficiency gains of these two groups of participants. Only one significant correlation was found between the dependent variable and informal L2 contact: speaking French with American classmates was negatively correlated with OPI gains [r=-.386, p=.046].

2.1.2.2b *Location of SAE programme*

Dewey, Bown, Baker, Martinsen, Gold and Eggett (2014) state that the vast majority of empirical research into SAE 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 to this tendency because their sample compared the informal L2 contact of study abroad learners in Spain, Mexico, France, Russia, China and Egypt. Regression analysis revealed that studying in Mexico and Egypt was predictive of higher overall (i.e. formal and informal combined) L2 contact and of informal L2 contact [R^2=.103, $F(2,113)$=6.40, p=.002]. However, both of these SAE programmes provided organised opportunities for informal L2 contact: 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 these interactions in regular interviews with programme leaders. Thus, it was not the location of the SAE *per se* that impacted the L2 contact, but rather features of the SAE programme in those particular locations.

In a qualitative approach to location of SAE, Dewey, Ring, Gardner and Belnap's (2013) study of the social networks of n=71 SAE learners of L2 Arabic in Jordan and Egypt found that location was one of the top two influencing factors on the formation of social networks: participants in Jordan were housed close to a university campus and as such had access to native L2 speakers of their own age, and the impact of this was broader and deeper relationships with native speakers. Participants in Egypt lived close to shops, so the majority of their native speaker interactions were short, repetitive and less meaningful. In their huge quantitative study of military students on study abroad programmes globally, Watson, Siska and Wolfel (2013) found that only their participants who studied abroad in

China and Western Europe made significant gains in terms of regional knowledge. As there were no controls for living arrangements or language contact in this study, we are once more unable to determine whether this outcome difference was in fact a result of different SAE location.

2.1.2.2c *Length of stay*

How much time in the target language community is sufficient for any positive impact on L2 proficiency to take effect? Given the relative affordability and convenience of a shorter stay 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 who enrolled in a SAE 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 in his article 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 an academic SAE of either two or three months. Pre and post-SAE 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 SAE 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 SAE on the pragmatic development of four length of stay sample groupings.

There is also evidence of an impact of length of stay on vocabulary development: Ife, Vives Boix and Meara (2000), in their longitudinal study of the impact of SAE 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 finding should not be generalised without caution.

2.1.2.3 Language contact

2.1.2.3a *Measuring language contact*

Informal language contact (i.e. contact which occurs outside of the classroom) has been measured most frequently in study abroad research using a questionnaire called the Language Contact Profile, henceforth referred to as the LCP. The LCP was developed by Freed, Dewey, Segalowitz and Halter (2004) for use with L1 English learners of L2 Spanish. The original instrument comprises two sections; one for use before the SAE and the other to be administered after the SAE has ended. The pre-SAE section collects demographic data (e.g. gender; age) and information about the respondent's proficiency in the L2, including the number of years of formal L2 instruction and previous experience of living in the target language community. There then follow 13 items that describe contact with the L2 in the home country, e.g. "Reading Spanish language newspapers" (2004:352), to which the respondent chooses the most appropriate answer from a 5-point frequency scale (from 'almost never' to 'daily'). The post-SAE section of the LCP collects data about the respondents' living arrangements during the SAE, including whether they lived with L1 and/or L2 speakers, and then presents 41 items that describe contact with the L2 and the L1. The respondent must state on average how many days per week (from zero to seven) during the SAE they engaged in each listed type of contact, and then on average how many hours per day (from zero to more than five). The items are listed in order of skill; in other words, all of the types of contact that involve speaking in the L2 are listed

together, followed by reading, listening and writing, and then the same structure is repeated for contact scenarios that involve the L1.

It is likely that the LCP has enjoyed wide use in study abroad research due to its comprehensive coverage of the types of informal L2 contact a study abroad learner might encounter, and also due to a lack of available alternatives. For such a widely used instrument, however, there is little evidence as to its validity and reliability. Martinsen (2011) adapted the LCP for use in a study of the impact of a short-term SAE on cultural sensitivity in American learners of Spanish and reports a strong Cronbach's alpha coefficient of .83, yet the researcher failed to include in his article any specific details regarding how the instrument was adapted and changed from Freed *et al*.'s (2004) original. Briggs (2014) adapted the LCP for use in a study of the development of a research tool called the Opportunities with Language Simulator (OWLS), which comprises computer-based simulations of informal L2 contact scenarios. Sampling n=95 SAE learners in the UK who comprised a range of L1s and nationalities, she used her adaptation of the LCP to define which L2 contact scenarios were most identified with by the participants in order to decide which informal contact scenarios to simulate in the OWLS. Briggs (2014) used a 5-point 'how true of me' scale in her version of the LCP 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, she amended the wording of questionnaire items in the original instrument that began with "I try to..." so that the items represented scenarios encountered 'naturally' by SAE learners as opposed to strategies they employed to increase their L2 contact. Similarly to Martinsen (2011), Briggs (2014) reported a high Cronbach alpha coefficient of internal consistency (0.85) with reference to her adaptation of the LCP.

A group of study abroad researchers (Dewey, Bown & Eggett, 2012; Dewey, Ring,

Gardner & Belnap, 2013; Dewey, Bown, Baker, Martinsen, Gold & Eggett, 2014) have recently turned their attention toward investigating the social networks of SAE learners as a different approach from measuring the amount of L2 contact SAE learners encounter. The Study Abroad Social Network Questionnaire (SASIQ) is a nine-item questionnaire that was designed to measure the social networks of learners in terms of their interaction with native speakers of the L2 during a SAE. The authorship of the SASIQ is at the moment somewhat opaque: the first use of this instrument was reported in Dewey, Belnap and Hillstrom (2013). In Dewey, Bown & Eggett (2012:120) it is stated that the instrument "was developed by the first author" yet no date for this development is provided. In Dewey, Ring, Gardner and Belnap (2013:272) the instrument is attributed to Dewey *et al*. (2012). The SASIQ was based for the most part on Segalowitz and Ryder's (2006) Montréal Index of Linguistic Integration, which is a self-report measure of contact with both the L1 and the L2. The SASIQ is designed to measure SAE learners' social networks in five domains (Scott, 2000): (1) size, whereby the number of native L2 speakers a learner associates with is noted; (2) intensity, pertaining to the strength of each of the native-speaker associations and measured by asking the learner to rate on an 8-point scale each association from 'acquaintance' to 'very close friend'; (3) durability, which refers to the frequency of contact with each native-speaker association, regardless of whether that contact is in the L1 or the L2; (4) density, which pertains to the extent to which there are connections between a learner's native-speaker associations (Knoke & Yang, 2008; Scott, 2000), whereby the learner subdivides their native-speaker associations into friendship groups and subsequently the mean size and largest size of their social groups is calculated; and finally (5) dispersion, which represents a quantification of the different social groups in which the learner participates (e.g. book clubs, coffee mornings). Unfortunately, it is not possible first hand to evaluate the SASIQ because the instrument itself has so far not been

published: in Dewey, Belnap and Hillstrom (2013:90) the authors direct the reader to Appendix A after their first mention of the instrument, yet there are no appendices to the article.

2.1.2.3b *Types of language contact in SAE*

What types of L2 contact do language learners encounter on a SAE? In a mixed methods study of language contact on a SAE, Coleman and Chafer (2010) approached this question from the perspective of how the amount of L1 contact on a SAE might inhibit their participants' take-up of any available L2 contact. They administered a questionnaire to n=42 L1 English L2 learners of French who had studied abroad for one or two academic terms at the University of Dakar, Senegal, between 1988 and 2010 as part of an undergraduate degree in French Studies at Portsmouth University. The researchers were interested in discovering the extent to which contact with the UK via telephone and Internet may have prevented the sample from taking advantage of the L2 contact opportunities available to them in Senegal. The questionnaire required respondents to state on a 4-point scale from 'monthly at most' to 'daily' the frequency with which they had contacted the UK by phone and by Internet during their sojourn abroad. Coleman and Chafer (2010:156) state that the questionnaire was devised based upon "existing knowledge and, where available, on existing instruments", but they do not cite either. They found that, as one might expect, the frequency of contact with home via phone and Internet by their participants increased in line with the wider availability and lower cost of using these technologies over time: that is, those participants who studied in Senegal more recently reported more frequent use of phone and Internet to contact the UK, and most of them did so at no cost (e.g. using Skype or FaceTime). The qualitative data revealed that a number of participants who had regular contact with the UK during their SAE felt that this contact inhibited their contact with the L2. For example, one participant said that being in

a long-distance relationship with a partner in the UK who was often available to talk online acted as 'comfort blanket' to soothe feelings of isolation or homesickness, which may otherwise have spurred attempts to integrate into the target language community.

Hernández (2010) sampled n=20 L1 English L2 Spanish learners on a semester-long SAE in Madrid to investigate the effects of motivation and L2 contact on oral proficiency. He administered a modified, 10-item version of the LCP (Freed *et al*., 2004) after the SAE as a measure of language contact. The LCP data suggested that the sample spent the most time in scenarios that involved speaking Spanish to native or fluent speakers of Spanish, writing homework assignments in Spanish and listening to music in Spanish. The scenarios they reported the least amount of time in were listening to Spanish movies and videos, writing emails in Spanish, and reading Spanish magazines. Regression analysis revealed that higher integrative motivation was a significant predictor of "interaction with the L2 culture" (2010:606) as measured by the LCP, which actually measures amount of L2 contact. Integrative motivation explained 46% of the variance in the L2 contact data [β=.667, t=3.660, p=.002].

Briggs (2014) administered her modified version of the LCP to n=95 learners of English on a SAE in Oxford to determine which L2 contact scenarios her sample identified with most highly. She conducted exploratory principal component analysis on the questionnaire data to find three underlying components: the first comprised solitary L2 contact scenarios (e.g. reading a newspaper), labelled 'Individual'; the second component comprised scenarios that involved interaction with others (e.g. short exchanges with cohabitants), and was therefore labelled 'Interactive'; and the third component involved scenarios in which the learner received information in English for purposes of either entertainment or practical necessity (e.g. watching films or asking for directions), and was labelled 'Informative'. The Informative component comprised the most scenarios and the most

heavily weighted scenarios. Briggs (2014) suggests that her sample most highly identified with Informative L2 contact scenarios because, unlike in ESL and EFL settings, in the SAE context the learner is often reliant on the L2 in the home setting; that is to say, ESL and EFL learners are more likely to live with L1 speakers and as such are less likely to be reliant on the L2 to mediate their relaxation time.

2.1.2.3c *Language contact and linguistic gain*

There is a wealth of evidence to suggest a positive relationship between L2 contact and linguistic gain. In terms of oral fluency, Yager (1998) studied the development of n=30 learners of Spanish on a summer abroad in Mexico. Pre and post the experimental period he administered an oral proficiency test (comprising a picture description task) and an early version of the LCP, which was based on Day (1985) and Seliger (1977). He found that interactive L2 contact was positively correlated [r=.39, p<.05] with native-like speaking overall, whereas solitary L2 contact was negatively correlated with gains in native-like pronunciation [r=-.41, p<.05].

In another study of L2 contact and oral proficiency gain, Freed, Segalowitz and Dewey (2004) compared three groups of L2 learners of French (in SAE, EFL and immersion settings respectively) in terms of their contact with the L2 and oral fluency. They found counter-intuitively that the SAE group neither experienced the greatest amount of L2 contact, nor saw the greatest gains with regard to fluency. Segalowitz and Freed (2004) studied a group of n=47 L1 English L2 Spanish learners comprised of n=18 AH learners in Colorado, U.S.A. and n=22 SAE learners in Alicante, Spain over a period of one academic term. A number of measures were administered pre and post the study period: (1) the original LCP (Freed *et al*., 2004); (2) the ACTFL Oral Proficiency Interview (OPI); (3) a computer-based lexical access test, which involved respondents deciding whether nouns in

both the L1 and L2 were living or non living things. This test measured speed and accuracy of responses; (4) an attention control test, which involved respondents deciding which word from a choice of three was matched with or did not match a stimulus word. This fourth test yielded data on speed and efficiency of attention control in both the L1 and L2. The findings suggested that the SAE group made greater gains in oral proficiency than did the AH group, but there was no significant relationship between amount of L2 contact and gains on the OPI. There was no significant relationship between lexical access and L2 contact, nor between attention control and L2 contact. However, for the SAE group longer mean length of turn at the pre-test was positively correlated with higher reports of L2 listening, and more hesitation-free oral production at pre-test was positively correlated with L2 contact that involved reading.

Dewey (2008) compared the vocabulary growth of three groups of English learners of L2 Japanese: at home (n=22), domestic immersion (n=14) and SAE (n=20). He administered three vocabulary measures: a picture matching test, a situational vocabulary test and an adaptation of the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993). To measure amount and type of L2 contact Dewey (2008) administered the LCP (Freed *et al*., 2004) and also employed weekly learner journals. The data on informal contact gleaned from these instruments was used to ascertain whether a correlation existed between informal L2 contact and lexical growth. The SAE and domestic immersion participants outperformed the at home group on picture matching and depth of vocabulary knowledge (as measured by the Vocabulary Knowledge Scale), and the SAE group outperformed both other groups on situational vocabulary knowledge. In terms of L2 contact, Dewey (2008) found a significant positive correlation between time spent speaking Japanese on the SAE (with friends or a host family) and gains in situational vocabulary knowledge. A significant

negative correlation was found between reading on the Internet in the L2 (e.g. surfing the web and reading emails) and gains in depth of vocabulary knowledge.

Hernández (2010) investigated the effect of motivation and L2 contact on oral proficiency. His sample comprised n=20 L1 English learners of L2 Spanish who enrolled in a semester-long SAE in Madrid. Two measures were administered pre and post the sojourn abroad: (1) the Simulated Oral Proficiency Interview (SOPI) as a measure of oral proficiency development; and (2) a questionnaire which collected background and demographic data, and which also included a motivation index that measured participants' integrative and instrumental motivation for learning Spanish. Additionally, a modified, 10-item version of the LCP (Freed *et al*., 2004) was administered after the SAE as a measure of language contact. There was a significant gain in oral proficiency from pre to post-test [t=–6.842, df=19, p=.000]. Amount of L2 contact was a significant predictor of oral proficiency gains [β=.693, t=4.080, p=.001], explaining nearly 50% of the variance in the gain scores.

Dewey, Bown and Eggett (2012) employed an online version of the LCP, which they had modified in response to much of the critique of the original instrument, to investigate the language contact of n=204 learners of Japanese in Japan. Firstly, whereas in the original instrument respondents had been asked to report how many hours per day and how many hours per week they spent in a particular kind of L2 contact, the modified version required only an estimate of the number of hours per week. The computer-based survey then generated an estimated number of hours per week that each learner spent engaged in speaking, listening, reading and writing in the L1 and L2 respectively. Respondents were then given the opportunity to "consider the feasibility" (2012:120) of these calculations. The other major modification the authors made to the LCP was to begin the questionnaire by asking respondents to estimate how many hours a week they spent on listening to, reading, writing and speaking in the L1 and L2 respectively, before going on to ask more

specific questions about how much time was spent in particular types of contact within those broad skills-based categories. The researchers made this modification in response to concerns about over-inflation of responses to the original instrument.

With regard to language contact and phonological control, Muñoz and Llanes (2014), in their study of the impact of age, learning context (AH or SAE) and L2 contact on the degree of foreign accent of Spanish learners of L2 English, administered a questionnaire that was remarkably similar to the LCP (Freed *et al*., 2004) to find that gains in degree of foreign accent were positively correlated with overall number of hours spent speaking English (p=.003), number of hours of speaking English in class (p=.002), and number of hours of speaking English to native speakers (p=.017).

There is evidence of an impact of language contact on intercultural competence: Martinsen (2011) sampled n=45 L1 English, L2 Spanish university students in the U.S. who had enrolled in a six-week SAE programme in Argentina. He investigated the impact of four independent variables on development of his sample's cultural sensitivity. The variables were pre-SAE motivational intensity, pre-SAE oral skills, relationship with host family and interaction with native Spanish speakers. Martinsen (2011:127) "used an updated version of the LCP" to measure language contact but does not report how this version was different from Freed *et al.*'s (2004) original. He administered the LCP after the SAE and as such his findings with reference to language contact may have been confounded by reliance on the participants' memory. Moreover, the LCP is not a measure of interaction with native speakers but rather a measure of time spent in informal contact with the L2, whether solitary or interactive. Regression analysis revealed that 'interaction with native L2 speakers' (i.e. informal L2 contact) was the only variable significantly to predict gains in cultural sensitivity (p=.01). However, the relationship between informal contact and cultural sensitivity gain was curvilinear, suggesting that the more informal L2 contact a

participant had, the more their cultural sensitivity increased, but that this trend was only true up to a certain point: the participants who reported the very highest amounts of informal contact gained less cultural sensitivity than those who reported slightly less informal contact. In explanation of this finding Martinsen (2011:131) suggested that some learners may engage in too much informal L2 contact, which leaves them fatigued and without time to 'digest' their experiences.

In a study of the relationships between social networks, L2 contact and linguistic gain in an Arabic study abroad context, Dewey, Belnap and Hillstrom (2013) used the SASIQ and the LCP as their independent variable measures. They sampled n=30 L2 learners of Arabic, who were studying either in Jordan or Morocco, and used a 'then-and-now' self-report survey based on Clark (1981) as their measure of proficiency. The researchers fail to report the L1 of the learners or the length or nature of the SAE. The proficiency survey required respondents to state on a 5-point scale from 'not at all' to 'quite easily' how able they were to perform different tasks in the L2 such as, "Give simple biographical information about self (name, age, composition of family, etc.)" (2013:96). This measure was administered before and after the SAE. The researchers argue that a self-report measure is an appropriate measure of proficiency because "self-assessment has frequently been used in study abroad research" (2013:91) and thus they assert that their findings from using this measure are comparable to other study abroad research. However, the SAE studies they cite in support of this claim were all conducted in the 1990s, which were the very early days of study abroad research. Certainly, the vast majority of quantitative research in this field at present makes use of objective measures of language acquisition.

In Dewey, Belnap and Hillstrom's (2013) study a web-based version of the LCP, which sounds very similar to the adaptation used in Dewey, Bown and Eggett (2012), and the SASIQ were administered after the SAE. The LCP in this case yielded a total number of

hours spent engaged in each of the four skills in the L2. The researchers state that the purpose of calculating these figures was not to estimate how much time was spent in engaged in each skill, but rather to know which kinds of L2 contact were more or less frequently occurring: they argue that if more time was reported in one skill, then L2 contact using that skill must have been more frequent than a skill for which was reported less time. The flaw in this argument is clear: if a learner engaged in one long conversation, the amount of time reported would be the same as for many short verbal exchanges, and thus an estimation of time spent is by no means a valid measure of frequency. *T*-tests indicated that scores on the post-SAE proficiency measure were significantly higher than the pre-SAE survey scores ($p<.0001$). Regression analysis was conducted to determine the ability of the participants' social networks and language contact to predict their gain in self-reported proficiency. The model explained 80% of the variance in proficiency gain [$R^2=.808$, $F(3,15)=26.2$, $p<.0001$]. The researchers report three significant predictor variables, all of which pertained to social networks: having Arab friends who had a high level of proficiency in English, having stronger relationships with Arab friends, and interacting with people outside of their social network were all predictive of L2 gains.

2.1.2.3d *Obstacles to language contact in SAE*

The literature suggests that SAE learners experience less informal interactive contact with native speakers during a SAE than they might have expected (DeKeyser, 1986; Polanyi, 1995; Rivers, 1998; Wilkinson, 1998). Some of the variables that can inhibit or modify a learner's contact with the L2 have already been discussed (e.g. gender, proficiency, age). Living arrangements have also been shown to play a key role in language contact. For example, and as stated by Briggs (2014), language learners who live by themselves during the SAE, 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 SAE

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 SAE learner could stem from disinterest or emotional distancing on the part of the host family. Certainly, the promise of supportive L2 interaction suggested by a stay with a host family has been shown not always to materialise (Wilkinson, 1998; Twombly, 1995). Other qualitative studies have proffered different reasons for non-engagement in language contact on a SAE. For example, Pellegrino Aveni's (2007) exploration of the experiences of U.S. language learners on a SAE 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. In summary, whilst there is an overwhelming amount of empirical evidence in support of the positive impact of L2 contact during a SAE on linguistic outcomes, there is also evidence to suggest that not all learners are offered the same opportunities. Moreover, learners may for a range of reasons choose not to take advantage of the opportunities which do present themselves.

2.1.3 SUMMARY OF STUDY ABROAD EXPERIENCE

As we have seen, the current evidence on study abroad experience and second language acquisition points to a strong impact of the former on the latter with regard to gains in oral proficiency and vocabulary knowledge, yet in terms of writing, listening, reading, grammatical competence and phonological competence the picture is less clear. The literature suggests a positive impact of study abroad experience on the self-oriented beliefs of second language learners but also a possible negative impact on perceptions of study abroad and of the target language culture. A number of variables have been found to mediate the impact on linguistic gain of a period abroad, and the most convincing evidence

points to a strong influential role of initial proficiency level, length of stay and language contact.

2.2 VOCABULARY KNOWLEDGE

In this section the following areas of second language vocabulary knowledge are explored: the relationship between vocabulary knowledge and second language competence; the nature of word knowledge, both in terms of single words and multi-word lexical items; processes of second language vocabulary acquisition; and assessing vocabulary knowledge.

2.2.1 VOCABULARY KNOWLEDGE AND SECOND LANGUAGE ACQUISITION

Scholars have widely cited vocabulary knowledge as a key contributor to second language ability (e.g. Wilkins, 1972; Nation, 2001; Milton, 2013). Laufer (1989) estimated that a learner needs to know approximately 95% of the lexical items in a language in order sufficiently to understand it, referred to as *lexical coverage* (Adolphs & Schmitt, 2003) and this figure has been refined more recently to stand at 98-99% with reference to written English (Hu & Nation, 2000), which amounts to only one unknown word in every 50 (Schmitt, 2008). For both written and spoken English, studies have reached the conclusion that anything less than 90% lexical coverage results in poor comprehension (e.g. Bonk, 2000; Hu & Nation, 2000). As stated by van Zeeland and Schmitt (2013), estimations of the lexical coverage required adequately to comprehend L2 texts are important because they provide an estimation of the vocabulary size a learner requires.

With regard to the value of vocabulary knowledge in consideration of grammatical knowledge, Wilkins (1972:111) said, “While without grammar very little can be conveyed, without vocabulary nothing can be conveyed”. A number of theories of language learning

derive from the belief that acquisition of grammatical features is driven by and dependent upon vocabulary knowledge. For example, Pienemann's (1998) *Processability Theory* contends that activation of lemmas (the canonical form of a word its inflections) in the mental lexicon is the first process in an incremental hierarchy of processes that result in language acquisition, and as such, without knowledge of lemmas to be activated (lexical knowledge), language learning cannot take place. Chomsky's (1957) theory of *Universal Grammar* holds that certain grammatical parameters are innate, and building on this concept the *Lexical Learning Hypothesis* (Ellis, 1997) posits that lexical learning drives a restructuring of the innate grammar. Certainly, there is evidence to suggest that grammatical development occurs as a result of lexical learning, particularly with reference to learning multi-word lexical items (Myles, Hooper & Mitchell, 1998; Bardovi-Harlig, 2002).

As stated by Milton (2013:71), "The single factor of vocabulary can explain up to 50% of the variance in performance in scores gained from tests of the four skills." Vocabulary size has been found to correlate positively and significantly with reading comprehension (Laufer, 1992; Qian, 1999; Albrechtsen, Haastrup & Henriksen, 2008; Stæhr, 2008). Moreover, there is evidence to suggest that vocabulary size and automaticity (that is, how accessible the word knowledge is) are positively correlated with reading and writing scores (Schoonen, 2010). There is 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), and vocabulary size has been found to be equally predictive of oral and written skills in English (Milton, Wade & Hopkins, 2010).

Recent research suggests that over and above its relationship to grammar and the four skills, vocabulary size may in fact impact on other areas of linguistic competence. For example, Bundgaard-Nielsen, Best and Tyler (2011a) studied n=31 adult learners of

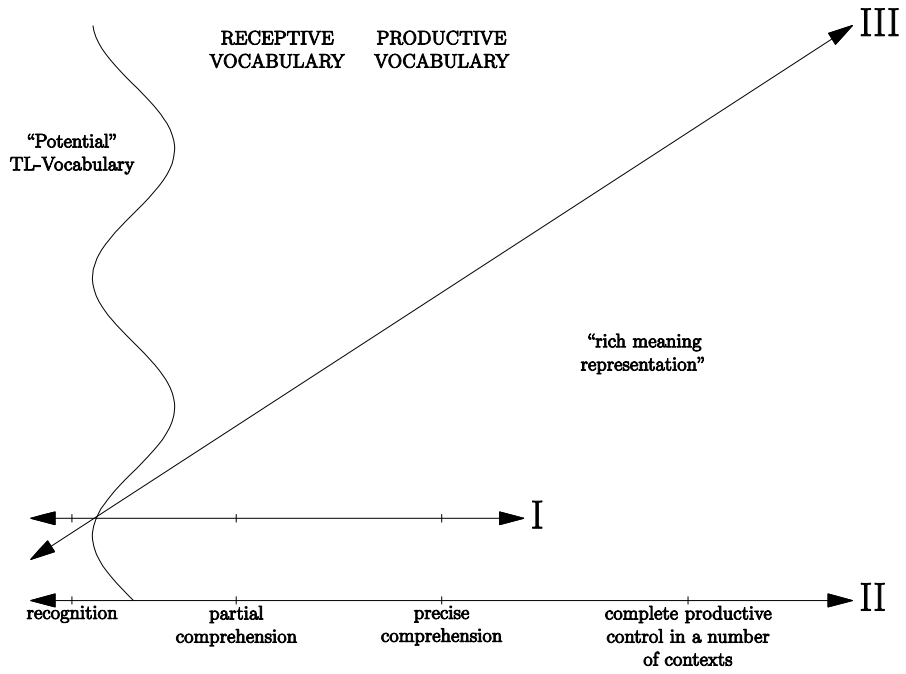
English in Australia to find that their participants' vocabulary size was positively correlated with their phonological inventories. That is, those participants who were able more accurately to perceive and discriminate between L2 vowel sounds were those with a larger vocabulary range. Additionally, Bundgaard-Nielsen *et al.*'s (2011b) longitudinal study found that development in perception and discrimination ability was linked to development in vocabulary size, as opposed to being an effect of L2 exposure during the SAE.

2.2.2 KNOWING A WORD

What does it mean to 'know' a word? In addressing this question a number of scholars have proposed frameworks of vocabulary knowledge to which the literature refers. Richards (1976) proposed a seven-part framework that comprised reference to knowledge of syntactic and semantic information, word associations and form. Chapelle (1998) argued for the consideration of four dimensions of vocabulary knowledge: (1) size or breadth of word knowledge; (2) knowledge of the characteristics of individual words; (3) the organisational structure of the lexicon; and (4) processes governing lexical access. Meara (1996) posited a bi-dimensional model of lexical knowledge that comprised consideration of size and organisation. Henriksen (1996; 1999) extended Meara's model by devising a model of vocabulary acquisition that divided word knowledge into three continua of lexical control (as shown in figure 2.1): (1) extent of knowledge (from partial to exact); (2) aspect of knowledge (receptive to productive); and (3) depth of knowledge (shallow to deep). However, whether these arenas actually represent continua or are dichotomous is under debate (Laufer & Goldstein, 2004). Many researchers, similarly to Henriksen (1996; 1999), have argued that receptive and productive word knowledge are different levels of the same type of knowledge (Melka Teichroew, 1982; Færch, Haastrup & Phillipson, 1984; Palmberg, 1987; Tréville, 1988; Read, 2000). However, Meara (1997)

contends that these types of word knowledge may operate differently; that is, that receptively known words are connected unidimensionally to items in a learner's mental lexicon whereas productively known words have different kinds of links to the lexicon, and as such productively known words can be activated when other words are triggered. Schmitt (2010:82) thus argued that to truly measure the depth of knowledge of a productively known word would entail calculating the number and nature of connections by which it was linked to existing items in the lexicon.

Figure 2.1: Henriksen's (1996) model of vocabulary acquisition



I: Partial to precise knowledge; II: Receptive to productive knowledge; III: Shallow to deep knowledge

Nation (2001:29) proposed a nine-part taxonomy to classify the multidimensional nature of word knowledge (shown in table 2.1). The nine subdivisions were grouped into three categories: (1) form, which comprises receptive and productive knowledge of the spoken, written and word part forms of a lexical item; (2) meaning, which pertains to receptive and productive knowledge of the form-meaning connection, the concept(s) behind, the referents for and the meaning associations carried with a lexical item; and (3) use, which comprises the receptive and productive knowledge of the grammatical function(s),

collocations and restraints on usage allied to a particular lexical item (e.g. register, frequency level).

Table 2.1: Nation's (2001) taxonomy of word knowledge (adapted from Nation, 2001:27)

Form	Orthographic
	Phonological
	Word parts
Meaning	Form and meaning
	Concepts and referents
	Associations
Use	Grammatical function
	Collocations
	Constraints on use

Qian (2002) and Laufer and Goldstein (2004:401) proposed an additional aspect of lexical knowledge; that of automaticity (sometimes also referred to as fluency), which refers to the adequacy of speed with which a lexical item is comprehended or produced (Schmitt, 2010).

The three features common to all of the vocabulary knowledge frameworks here discussed are the inclusion of reference to (1) vocabulary size (i.e. the breadth of both receptive and productive word knowledge), (2) depth of word knowledge (the extent of receptive and productive knowledge of form, meaning and use), and (3) lexical access (automaticity of comprehension and production). Daller, Milton and Treffers-Daller (2007) refer to these three elements as the three dimensions of lexical space (illustrated in figure 2.2).

Figure 2.2: The three-dimensional lexical space (Daller *et al*., 2007)

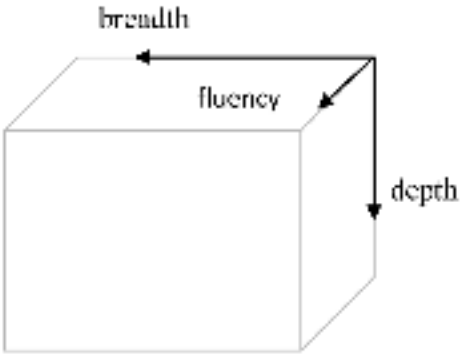


Illustration taken from Milton (2013:62).

2.2.3 KNOWING MULTI-WORD LEXICAL ITEMS

Disparity exists between academics as to a fixed definition of the term 'multi-word item', and as such a number of different nomenclatures and subcategories exist, such as formulaic language (McCarthy, 1990; Wray, 2002) or formulaic sequences (Schmitt, 2004; Wood, 2006). Furthermore, within the category of multi-word phenomena exists a host of sub-categories including collocations ('blonde hair'), phrasal verbs ('to get over something/someone'), idioms ('put all your eggs in one basket'), compounds ('peanut butter'), fixed/semi-fixed phrases ('What's up?'/ 'Pass the …') and prefabs (prefabricated utterances learned and produced as 'chunks' rather than being generated grammatically). One feature common to all operationalisations of these terms is the presence of two or more words that occur together frequently, often in a set sequence. For the purposes of the present study, the term *multi-word lexical item* will be used to encompass all of this type of language.

There is convincing empirical evidence to indicate that multi-word lexical items are stored in and retrieved from the lexicon as single 'chunks' (Schmitt, 2004; Wray, 2002; Wray, 2008). From a theoretical perspective, Cutting and Bock (1997) posit a hybrid model of how multi-word items are stored in the lexicon: they propose that each multi-word item

has its own entry in the lexicon which, when activated, in turn activates related lexical concepts (whether they are represented by individual or multi-word items). However, they also argue that when a multi-word item is activated the individual words that make up the item are also activated and in turn they activate their lemma connections. Thus, in Cutting and Bock's (1997) model, the unit 'over the hill' would activate conceptually related phrases (e.g. 'long in the tooth', 'one foot in the grave') and also singular words related to the original multi-word item (e.g. 'across', 'under', 'road', 'mountain').

There is evidence to suggest that multi-word lexical items represent a high percentage of spoken and written English (Nattinger & DeCarrico, 1992): Erman and Warren (2000) suggest that 58.6% of spoken English and 52.3% of written English is comprised of multi-word items, whereas Biber, Johansson, Leech, Conrad and Finegan (1999) posited that multi-word items comprise 30% of conversational English and 21% of academic writing in English. Altenburg (1990) has argued that up to 70% of English as used by adult native speakers can be classed as formulaic. In a corpus-based study, Martinez and Murphy (2011) found that almost 500 multi-word items exist in the first 3,000 most frequent word families used in English. In another corpus-based study, Shin and Nation (2008) found that 84 collocations occur in the 1,000 most frequently spoken English words and 224 in the second 1,000. Clearly, then, multi-word lexical items are an important and pervasive feature of the language, and they make a significant contribution to vocabulary size (Wolter & Gyllstad, 2013). As such, they are an important component of second language vocabulary for learners of English to acquire.

Despite their prominence in native speaker production, it has been argued that multi-word lexical items are difficult for second language learners of English to acquire (Barfield & Gyllstad, 2009; Laufer & Waldman, 2011). One reason posited for this is that second language learners, particularly those who do not live in the target language community (i.e.

EFL learners), do not have repeated and varied exposure to multi-word items sufficient to notice them as important contributors to the language (Durrant & Schmitt, 2010). English language course books and native speaker corpora have been compared to find that multi-word units are underrepresented in course materials (Koya, 2005). Additionally, it has been argued that in receptive language processing learners tend to focus on individual words over and above lexical chunks (Wray, 2002; Gyllstad, 2007; Barfield, 2009). It has also been claimed that less semantically transparent multi-word items (such as those with figurative meanings) are problematic for language learners because they are less straightforward to comprehend (Henriksen, 2013).

Multi-word items have been investigated from three different conceptual perspectives: (1) frequency-based (Firth, 1957; Halliday, 1961, 1966; Sinclair, 1987, 1991), which is concerned with text analysis and calculation of the probability of word combinations (Barfield & Gyllstad, 2009); (2) phraseological (Cowie, 1981, 1988, 1998; Howarth, 1996, 1998; Nesselhauf, 2005), which analyses word combinations themselves in terms of semantics and syntax (Barfield & Gyllstad, 2009); and (3) generative, in which multi-word items are not generated as a composite of other words but rather produced as a single-word item would be; as a whole (Howarth, 1998). Moreover, Nesselhauf (2003) and Howarth (1998) highlight that the generative definition of multi-word items is limited only to those combinations of words that are used figuratively (i.e. differentiating between ‘eat lunch’ and ‘eat your words’).

2.2.4 VOCABULARY ACQUISITION

Vocabulary acquisition processes can be dichotomised as either resulting from implicit or explicit learning. As stated by Schmitt and McCarthy (1997), there has long been disparity over which of these two approaches to vocabulary acquisition is more beneficial.

2.2.4.1 Implicit vocabulary acquisition

Implicit vocabulary learning (also termed 'incidental' and 'unattended') refers to the unconscious acquisition of lexical items as a result of meaning-focused L2 exposure. Implicit vocabulary acquisition is relevant to the present study because one could argue that much of the informal L2 contact experienced on a SAE is meaning (as opposed to form) focused. Krashen's (1989) *Input Hypothesis* suggested that all language can be subconsciously acquired in this way, yet the concept of subconscious language acquisition becomes more complex in consideration of Schmidt's (2001) *Noticing Hypothesis*, which posits that to learn something, noticing (i.e. being consciously aware of it) is necessary. However, if noticing constitutes conscious awareness, acquisition ceases to be truly incidental because the form being acquired has entered conscious thought. As such, truly incidental acquisition may not in fact be possible. Ellis (1994) argued that it is possible for something that is on the periphery of conscious attention to be noticed, even where the main focus of attention is elsewhere, and that by extension it can remain incidental but noticed nonetheless. Implicit vocabulary acquisition is henceforth operationalised as the learning of lexical items from meaning-focused activity.

2.2.4.1a *Implicit vocabulary acquisition from reading*

Whilst implicit vocabulary acquisition is possible from listening (Vidal, 2003, 2011; van Zeeland & Schmitt, 2013), most research in this area involves reading because listening has been shown to be less effective as a platform for vocabulary gain (Brown, Waring & Donkaewbua, 2008; Vidal, 2011). There is much empirical evidence to show that incidental vocabulary acquisition is possible from reading (Day, Omura & Hiramatsu 1991; Hulstijn 1992; Dupuy & Krashen, 1993; Waring & Takaki, 2003; Pigada & Schmitt, 2006). Indeed, Ellis (1995) considers reading to be the ideal medium for this endeavour.

Nation (2001) agrees, stating that texts with a strong theme, which enable the repetition in a range of contexts of unknown words, provide the most effective input in terms of acquiring words incidentally. Day *et al.* (1991) sampled n=191 Japanese learners of English to test their ability incidentally to acquire seventeen target items from a reading passage of 1,032 tokens, which had been modified to increase the frequency of the target words and enhance the context for ease of lexical inferencing (i.e. guessing the meaning of the target words from the textual context). The two treatment groups (one each from high school and university students respectively) read the passage and then immediately took a multiple-choice format test to test their knowledge of the target lexical items, whereas the control groups took the test without having read the passage. The treatment groups significantly outperformed the control groups on knowledge of the seventeen items ($p<.01$ in both cases). This finding provided robust evidence in support of the existence of incidental vocabulary acquisition. However, as the measure used tested only receptive knowledge of the form-meaning link of the target lexical items, and as no delayed post-test was administered, the extent of the acquisition that took place (in terms of depth and longevity) is unclear. Moreover, the use of multiple-choice format testing means that guessing may have inflated the gain scores (Waring & Takaki, 2003).

Dupuy and Krashen (1993), in a similar methodological design to Day *et al.* (1991), used a 15-page reading passage to test the incidental acquisition by L2 learners of French of 30 target words. They found the participants had a mean gain of 6.6 words, yet how this figure relates to Day *et al.*'s (1991) findings is unclear because we are not told the number of tokens in the passage. Moreover, in Dupuy and Krashen's (1993) study two factors may have confounded the reported gains: Firstly, the sample watched a supplementary video in addition to reading the text, which may mean that the participants' focus during the reading task was less tied to comprehension of meaning and more on explicit consideration

of the language it comprised. Secondly, the sample was forewarned that they would take a vocabulary test after the reading task. As such, one could argue that their attention to the lexical items in the passage may have increased. Hulstijn (1992) extended this study design by incorporating a measure of retained comprehension; that is, he compared the retention of inferred meanings with given meanings from vocabulary glosses. Instead of using a multiple-choice test format, he asked his sample to state the meaning of the twelve target words. Hulstijn's (1992) findings suggested that in terms of retention, inferred meaning was better than given meaning. To justify this finding he proposed the *Mental Effort Hypothesis* (1992:113), based on the *Levels of Processing Hypothesis* (Craik & Lockhart, 1972), in which enhanced retention is explained by the requirement of deeper cognitive processes to infer a meaning than to read and comprehend one. However, not all of his participants' lexical inferences were correct, which suggests that whilst incidental vocabulary acquisition from reading is possible, without modification of the input it may not be effective. Moreover, the gains reported were low, lower even than those found by Dupuy and Krashen (1993).

More recent research into incidental vocabulary acquisition has focused on which dimensions of word knowledge are incidentally acquired and the relationship between incidental acquisition and the frequency of occurrences of a lexical item in a given text. For example, Waring and Takaki (2003) investigated the incidental acquisition of lexical meaning and form of 26 target words from a graded reader by n=15 Japanese learners of English. Although the incidentally acquired words had for the most part been forgotten by the three-month delayed post-test, the sample made small gains at the immediate post-test, and greater gains were made on recall of meaning than recognition of form. Fewer encounters with a word were needed to facilitate recognition of orthographic form and prompted recall of meaning than to facilitate unprompted recall of meaning. Moreover,

and perhaps unsurprisingly, the researchers found that those words most frequently encountered in the text were most resistant to attrition. In a similar study, Pellicer-Sánchez and Schmitt (2010) explored the acquisition of recognition of written form (spelling), recall of word form and recognition and recall of word meaning from reading a novel. The sample made gains on 28% of the 34 target words. The data was analysed with reference to the frequency of occurrence of target items in the text in two frequency bands to find that, conversely to Waring and Takaki's (2003) findings, fewer exposures to a word were required to facilitate word meaning recognition than recognition of the word form.

2.2.4.1b *Implicit vocabulary acquisition from listening*

Vidal (2003) sampled n=116 university students to investigate the acquisition of lexis from listening to academic lectures. The sample made significant gains as compared to pre-study vocabulary knowledge and those gains were still significant at the 4-week post-test. Vidal (2011) also compared the gains made from listening with those from reading to find that reading facilitates significantly higher vocabulary gain. Moreover, other than for learners of a very high proficiency level, Vidal (2011) found that vocabulary acquisition from reading is retained longer than words acquired from listening tasks. Van Zeeland and Schmitt (2013) explored the incidental acquisition of three dimensions of vocabulary knowledge (form, meaning and grammar) with n=30 learners of English, who were engaged in postgraduate study at a British university and had a variety of L1s. The sample listened to four thematically related passages, which had been treated to control for lexical coverage and frequency of occurrence of the 24 target words. Form and grammar recognition were measured using a multiple-choice format test. To measure meaning recall the target items were presented in the same sentence contexts as had been used in the grammar recognition test, and respondents were required to write down anything they knew about the meaning of each item. The tests were administered immediately after

listening to the four passages, and again two weeks later. The findings revealed that the sample did make vocabulary gains of up to 29% at the immediate post-test and 19% at the delayed post-test. Knowledge of form was acquired more easily than knowledge of meaning but by the delayed post-test even the gains in form and grammar were mostly lost. However, where meaning was acquired it was largely retained to the post-test. Unfortunately, as only n=10 participants completed the post-test the findings of this study with reference to retention should be treated with reserve.

With regard to the immediate post tests used in most of the studies of incidental vocabulary acquisition, research suggests that newly encountered lexis is not immediately entered into the mental lexicon but rather that a period of consolidation is first required (Gaskell & Dumay, 2003; Dumay & Gaskell 2005, 2007; Clay, Bowers, Davis, & Hanley, 2007). Gaskell and Dumay (2003:106) suggest therefore that before a lexical item has entered the lexicon, its recall or recognition is founded on an "episodic memory trace" related to its most recent (or frequent) occurrence. After the item has been added to the lexicon, its retrieval is activated by its links to other words and concepts in the mental network. By extension, immediate post-tests of vocabulary acquisition do not provide evidence of whether a word has been added to the lexicon, but merely provide an indication of the strength of the memory trace (Dumay, Gaskell, & Feng, 2004).

What does the empirical evidence tell us about incidental vocabulary acquisition? It is a complex process that is subject to a number of text-related factors. Certainly, the evidence suggests that vocabulary gain from listening and reading may not be retained unless a text has been modified to increase the frequency of occurrences of the target words. Also, one possible confounding variable applicable to all of the studies here discussed is that of context: one could argue that the linguistic and wider context in which a word occurs in a text has an impact on whether and to what extent it is noticed and/or acquired. Despite the

complexity, some patterns of evidence have emerged from the literature in terms of how incidental vocabulary acquisition develops: the acquisition pattern found by van Zeeland and Schmitt (2013) of form before grammar before meaning was also found by Waring and Takaki (2003), Webb (2007) and Chen and Truscott (2010), and these combined findings provide robust evidence in support of the incremental nature of vocabulary acquisition, and to suggest that recognition of form occurs with fewer exposure to a lexical item than recall of meaning.

2.2.4.2 Explicit vocabulary acquisition

Although incidental acquisition from reading (and to a lesser degree, listening) can result in significant vocabulary gain, naturalistic exposure is insufficient to acquire the size and depth of vocabulary knowledge required to operate successfully in the L2 (e.g. Cobb & Horst, 2004; Laufer, 2005; Ellis, 2008) and as such the argument for explicit approaches is strong (Hulstijn, 2003; Nation, 2007; Elgort & Nation, 2010). Explicit vocabulary acquisition constitutes the conscious attention of the learner to words in a text over and above their part in overall comprehension of meaning.

Swain (1985) posited the *Output Hypothesis* to account for her belief that "producing the target language may be the trigger that forces the learner to pay attention to the means of expression needed in order to successfully convey his or her own intended meaning" (1985:249). Swain (1993, 1995) has since elucidated on her hypothesis to characterise output as tripartite; comprising (1) hypothesis testing, in which the learner can judge the validity of their linguistic hypotheses based in the feedback they receive from interlocutors; (2) metalinguistic knowledge, whereby a learner engages in syntactic processing more deeply in output than in reception of input; and (3) noticing, whereby,

similarly to Schmidt's (2001) hypothesis, in producing the L2 the learner becomes aware of the gaps in their knowledge.

An extension of Swain's output hypothesis as put forward by Hulstijn and Laufer (2001), the *Involvement Load Hypothesis* posits that vocabulary acquisition is mediated by three factors: (1) need, which is the self or task-determined requirement for a word or its meaning); (2) search, where a learner finds a word for a concept or vice versa); and (3) evaluation, whereby the learner makes decisions about how a word is used in context. Hulstijn and Laufer (2001) tested the hypothesis by testing advanced proficiency EFL learners in the Netherlands and Israel on acquisition of ten target words as a result of a reading task and a modified output task. Group 1 read a text enhanced with marginal glosses. Group 2 read the same text that had 10 gaps (that could be completed with the target words). Group 3 were provided definitions and examples of the words and were then asked to write a letter using all ten. The immediate post-test required respondents to give translations of the target lexical items and to state whether they 'knew' these items before participation in the study. This is problematic because, as discussed in 2.1.2, there are many aspects of word knowledge and no operationalisation of the term was provided to the respondents. A more reliable method would have been to administer a pre-treatment vocabulary test. The researchers hypothesised that Group 3 would retain the most vocabulary due to the increased involvement load of the output task, and this was in fact the case. There was no significant difference between the retention of Groups 1 and 2. Although this finding provides some support for the Involvement Load Hypothesis, the differences between groups may have been a result of the differences in task types, and particularly with regard to task three because it involved an entirely different modality. To augment the argument in support of output as a means of consolidating vocabulary

acquisition, a similar study would need to be conducted in which the only difference between treatment groups was the involvement load.

In a study to determine the impact of differing amounts of exposure on explicit and implicit lexical acquisition, McLaughlin, Osterhout and Kim (2004) conducted an investigation into the word form knowledge of L1 English beginner level learners of L2 French and a control group who were not learning French. To test explicit knowledge of form the researchers administered a lexical decision task (real/pseudoword) and to test implicit knowledge of form they employed an *Event-Related Potential* (brain response) measure. The measures were administered at three time points: after 14 hours of French instruction, after 63 hours and after 138 hours respectively. The brain responses of the sample were measured whilst they were completing the lexical decision task. In terms of explicit acquisition of word forms, little improvement was made across the three testing sessions. However, the implicit knowledge measure revealed that even at the first testing session the sample had developed sensitivity to semantic relationships and lexical status (i.e. which pairs of words constituted a match and which were real as opposed to pseudowords). The researchers therefore suggest that explicit tests may not be sufficiently sensitive to measure implicit knowledge.

Sonbul and Schmitt (2013) conducted a study of n=35 native English speakers who were undergraduates at a UK university and n=43 non-natives who were postgraduates at the same university to test the differential effect of three learning conditions on explicit and implicit lexical learning. The learning conditions were: (1) enriched, whereby more frequent instances of the target lexis occurred; (2) enhanced; in which the target lexis was enhanced in some way (e.g. bold typeface, highlighted); and (3) decontextualized, whereby the lexis was presented without context and instruction was focused on its memorisation. For the native speakers, fifteen low frequency medical collocations were

chosen. For the non-natives, fifteen medical collocations were also chosen but in this case, the researchers tried to ensure that the participants would know each of the individual words in each collocation by ensuring that the lexical items all fell inside the first 3,000 lemmas of the BNC or the General Service List (West, 1953).

For both of Sonbul and Schmitt's (2013) sample groups, a control pair for each of the fifteen collocations was devised, which had the same second word and a very similar first word, but which did not appear as a collocation in the BNC (e.g. *vanishing lung* became *decaying lung*). A medical reading passage was selected that included ten of the intact collocations and it was modified thus: for the enriched condition, five of the collocations were embedded in the text; for the enhanced condition, five of the collocations were emboldened and in red font; and in the decontextualized condition five collocations were presented on a PowerPoint presentation, one collocation on each slide for ten seconds and each in bold, red font. The three learning conditions and three sets of five collocations were counterbalanced to form three experimental blocks to which the participants were randomly assigned. Two explicit and one implicit measures were employed immediately after the experimental phase and as delayed post-tests. The explicit measures were a cloze form recall test and a multiple-choice form recognition test. The implicit measure was a computerised priming task, which presented the respondents with the first word of one of the thirty collocations (intact and control) and required the participants to state (yes/no) whether the second word together with the first word constituted a real collocation.

The findings of Sonbul and Schmitt's (2013) study revealed that all three learning conditions had significant and durable impacts on both sample groups in terms of their explicit (recall and recognition) vocabulary acquisition, and in both cases the decontextualized condition had the greatest effect. However, the results with reference to implicit knowledge were less positive: for native speakers and non-native speakers there

was no effect of any of the conditions on implicit vocabulary gains. These findings suggest that explicit vocabulary knowledge does not necessarily imply implicit knowledge, and that small amounts of explicit exposure do not lead to implicit knowledge. However, it may also be the case that the measure of implicit knowledge was not sufficiently sensitive, and that gains may have been detected had a brain activity measure been employed instead, as in McLaughlin *et al.* (2004).

2.2.4.3 Summary of vocabulary acquisition

Why might meaning be more difficult to acquire implicitly than form? Laufer (2005), in making the case for explicit vocabulary instruction in second language teaching, set forth four reasons why implicit learning of vocabulary is not sufficient in terms of acquisition of meaning: (1) comprehension of meaning can be achieved without attention to individual words; (2) lexical inferencing is unreliable in terms of accuracy of comprehension; (3) if the meaning of a lexical item can easily be inferred, the depth of processing level required to do so may be insufficient to promote learning; and (4) retention is unlikely unless the learner re-encounters the newly learned word soon after first exposure to it. The literature on explicit vocabulary acquisition indeed suggests that explicit instruction leads to explicit lexical knowledge in both the short and longer term, but scholars have yet to determine the amount of explicit instruction necessary from which implicit knowledge can develop, and nor has the relationship (if one exists) between explicit and implicit knowledge been defined. In summary, therefore, the present evidence suggests that implicit *and* explicit approaches to vocabulary acquisition are necessary for a learner of English to gain the amount and type of lexical knowledge sufficient to operate in the L2.

2.2.5 TESTING VOCABULARY KNOWLEDGE

As stated by Milton (2013), it is not possible for an individual test to measure every aspect of vocabulary knowledge as set out in the frameworks and conceptual perspectives in 2.1.2. As such, testing has focused on particular dimensions of vocabulary knowledge: vocabulary tests for the most part have attempted to measure either the number of words that a learner knows (size or breadth) or how well they know them (depth). However, the vast majority of vocabulary tests operationalize the construct of vocabulary size as pertaining only to single-item lexical units. Moreover, in most tests of size and depth vocabulary is presented as a series of decontextualized, discrete items (Read, 2007) and as such vocabulary knowledge is operationalised as a trait (Read & Chapelle, 2001); that is, as knowledge which is stable, regardless of context.

2.2.5.1 Testing vocabulary size

As remarked by Pignot-Shahov (2012), in SLA research tests of vocabulary size are more frequently used than tests of depth, and their findings more widely accepted. This is likely due to the fact that size is a one-dimensional construct and as such vocabulary size tests can test a larger sample of lexical items. Consequently, estimates of vocabulary size can usefully be compared to achievement scores and can thus more easily be generalised to the wider second language learner population.

2.2.5.1a *The unit of measurement*

One of the key issues inherent in testing how many words a learner knows is that of defining what constitutes a word for the purpose of the test. The options include types, which refer to how many different words there are in a text; tokens, which are the total number of running words in a text; lemmas, which comprise the canonical form of a word

(e.g. sit) and inflections of that form (e.g. sat, sitting); and word families, which include all words that share the same headword (e.g. cry, cried, crying, decry, cry out, cry-baby). Nguyen and Nation (2011) argue that the use of word families as the unit of measurement is more appropriate for receptive vocabulary knowledge testing because in receptive knowledge testing (and dissimilarly to productive knowledge testing), if a learner recognises the headword, little extra knowledge is required for recognition of its inflections and derivational affixes. However, it has been argued (Bogaards, 2001) that no empirical evidence exists to support the assertion that knowledge of a headword allows for comprehension of all of its word family members. As argued by Schmitt (2010), the most useful of these to vocabulary testing are lemmas due to their equal applicability to receptive and productive knowledge, which in turn allows for comparison between the two.

2.2.5.1b *Frequency*

Corpus-based research has shed light on the value of different words to learners of English in terms of their frequency (Nation, 2006; Milton, 2009; Webb & Rodgers, 2009). The British National Corpus (BNC; Aston & Burnard, 1998) is a 100-million-word strong corpus, which has provided the basis for the development of word frequency lists on which a number of vocabulary size tests are based. Unfortunately, frequency profiling has so far only been conducted with reference to single word lexical items: Cobb (2013) argues for the inclusion of multi-word items in frequency profiling of the BNC and suggests that doing so would result in revised frequency counts and more accurate lexical profiling. Frequency lists have revealed that high-frequency words are for the most part shorter in length than their low frequency counterparts, and generally speaking they are applicable to a larger range of contexts (Pignot-Shahov, 2012).

Starting with Palmer (1917) and for over a century since, scholars have argued for a *Frequency Hypothesis*, which proposes that the more frequently a word occurs in a language, the sooner it will be acquired by a learner of that language as a result of repeated exposure to it. As stated by Milton (2013), the frequency hypothesis is important in the field of vocabulary testing because many of the most commonly used size tests, such as Nation's (1990) Vocabulary Level Test, are based upon this concept. There is empirical evidence in support of this hypothesis. For example, Milton (2006) sampled n=227 L1 Greek L2 learners of English, who comprised the full range of proficiency levels, and administered a receptive, orthographic yes/no test of vocabulary size which tested participants' knowledge of words from the first five 1,000 frequency levels. For each respondent a mean score for each frequency level was calculated. There was a statistically significant relationship between size of vocabulary and the scores on the frequency levels as revealed by ANOVA [F=93.727, p<.001], providing convincing support for the frequency hypothesis. Similar results were found by Richards and Malvern (2007) in a study of UK learners of L2 French, and by Aizawa (2006) in a study of n=363 L1 Japanese learners of English.

2.2.5.1c *Tests of vocabulary size*

The Yes/No test was developed by Meara and Buxton (1987) and was designed to measure receptive vocabulary knowledge; presumably knowledge of meaning. The test presents the test-taker with the orthographic form of a number of words (60% L2 words and 40% pseudo words) and the learner indicates whether they 'know' the words (yes) or not (no). The proportion of 'yes' responses to pseudo words adjusts the score derived from the 'yes' responses to real words and an estimation of the learner's vocabulary size can be arrived at. The X-Lex (Meara & Milton, 2003) is a more recent development of the yes/no format that tests knowledge of the first five 1000-word frequency bands. As in the Yes/No Test,

test-takers are presented with the orthographic form of real and pseudo words to which they must respond ‘yes’ or ‘no’, and the score is adjusted based on the number of ‘yes’ responses to pseudo words.

The P-Lex (Meara & Bell, 2001) is a measure of productive vocabulary knowledge in terms of the lexical richness of short written texts by lower proficiency learners. The text analysis does not calculate the number of words used per frequency level but rather yields a single score that represents the ratio of infrequent to frequent words used. The Y-Lex (Meara & Miralpeix, 2006) extends the capacity of the X-Lex by testing in the same format from the 5000 to the 10,000-word level. The Auralex (Milton, 2009) also uses the same format as the X-Lex, but presents the phonological as opposed to the orthographic form of the words. Therefore, when the X-Lex or Y-Lex and the Auralex tests are used in tandem, comparisons can be drawn between a learner’s knowledge of written and spoken word form (Milton & Hopkins, 2006; Milton & Riordan, 2006).

There are many problems inherent in the yes/no format for vocabulary testing. Firstly, what knowing a word constitutes for the purpose of the test is not specified for the respondents and as such one could argue that it is likely that different test-takers are reporting on different types and/or depth of knowledge, and therefore that the test does not measure a single construct. For example, one test-taker may interpret ‘knowledge’ of a word as being mere recognition of orthographic form whereas another may respond ‘Yes’ only to words they know how to use. Meara (1996:44) argued that recognition of a large number of words is indicative of more complex knowledge about those words (“A learner with a huge vocabulary and nothing else is a possibility, but something of a freak”), but there is no empirical evidence to support this assertion; in fact, studies have shown that recognition of form precedes acquisition of meaning and is strongly subject to attrition (c.f. 2.1.4.1c).

Another problem with the yes/no test format is that even where the score is adjusted based on the proportion of 'yes' responses to pseudo words, there is a strong chance that scores are inflated due to correct guessing in response to the real words tested. There is some evidence to suggest that respondents are wont to inflate their self-reported vocabulary knowledge on bi-optional tests (Chall & Dale; 1950; Janssens, 1999). Moreover, there is evidence to suggest that the yes/no format is not well suited to learners of lower proficiency levels due to a tendency on their part to claim that they know the pseudo words (Meara, 1996). It is also the case that the L1 of the test-taker plays a role in how they respond to the test. Meara and Buxton (1987) found that with French learners of English the cognate relationship between the L1 and L2 impacted upon responses to the pseudo words. L1 Arabic speakers are negatively impacted by the presentation of the target items in isolation because in Arabic, vowels are not written but rather furnished by the reader as a result of the textual context (Cobb, 2000).

Schmitt (2000:173) said that Nation's (1990) Vocabulary Levels Test is "the closest thing we have [to an] accepted standardised test of English vocabulary". The levels test was designed to reveal how many high-frequency words a learner knows and covers both receptive and productive vocabulary knowledge (the productive version having been developed by Laufer and Nation, 1999). Schmitt, Schmitt and Clapham (2001:62) state that "the levels test...should be seen as providing indication of whether examinees have an initial knowledge of the most frequent meaning sense of each word in the test".

The Vocabulary Levels Test was developed on the premise that learners of English are likely to know more high-frequency words than low-frequency words; the frequency hypothesis. It tests vocabulary knowledge at five frequency levels: 2000, 3000, 5000, Academic Word List (Coxhead, 2000) and 10,000. The words included in each level were established using word frequency data from a number of sources, including Thorndike and

Lorge (1944) and West (1953). The receptive test comprises 36 words and 18 definitions at each frequency level. Each question presents the test taker with 6 words and 3 definitions, so that 3 of the words function as distractors. The productive test (Laufer & Nation, 1999) comprises 18 words at each frequency level. Each question is a sentence in which the target word is incomplete. In both the receptive and productive parts, the use of nouns, verbs and adjectives as the target word is at a ratio of 3: 2: 1 respectively.

Although the levels test improves upon the design of the yes/no format in that it is not based on self-report but rather requires the test-taker to demonstrate their knowledge, the use of a multiple-choice format brings its own disadvantages; namely, that there is a chance that the correct answer will be guessed (in the format of the levels test this is a 1/6 probability). Comparisons between the scores of yes/no tests and the Vocabulary Levels Test have yielded conflicting findings: Mochida and Harrington (2006) found that test-takers tended toward overestimation of their knowledge in response to yes/no vocabulary checklists but not to multiple-choice tests such as the levels test, whereas Stubbe, Stewart and Pritchard (2010) found that respondents who took both types of test scored significantly higher on the levels test.

Gu (2014: personal communication) highlights a discord between the theoretical assumptions underpinning the Vocabulary Levels Test and the exposure to English most likely experienced by its respondents. That is, he states that the test is founded on the assumptions that (1) vocabulary is distributed statistically (i.e. Zipf's law), and (2) vocabulary acquisition is similarly distributed (i.e. that the most frequent vocabulary items are acquired before the less frequent items). As mentioned, the frequency levels in the Vocabulary Levels Test were derived from native-speaker corpora. Gu (2014) argues that, due to un-natural exposures to English, some L2 learners of English can remember a large quantity of low frequency words but lack a basic knowledge of high frequency words, and

therefore that as a vocabulary breadth measure the levels test is not entirely accurate. However, a large number of studies have used the levels test to find that it is well able to discriminate between learners of differing vocabulary sizes and that it is able to detect increases in size (e.g. Laufer, 1998; Laufer & Paribakht, 1998; Read, 1998; Schmitt, Schmitt & Clapham, 2001).

Webb (2008) states that on the receptive levels test, participants have a 17% chance of guessing an answer correctly, with that percentage rising depending upon the number of target words and distractors in each item that the respondent knows. However, one could argue by extension that guessing is an unavoidable problem in all monolingual tests of receptive vocabulary knowledge because for the test to be a test only of receptive knowledge, the respondent cannot be asked to produce language and must therefore be presented with options with which to match the target item. Despite the criticism levelled at the levels test, there is much evidence to support the assertion that it is a highly valid and reliable test of vocabulary size (c.f. 3.5.1.1 for an in-depth discussion).

The Lexical Frequency Profile (Laufer & Nation, 1995) purports to measure vocabulary size by analysis of a test-taker's free production in the L2: the proportion of frequent to infrequent vocabulary used is calculated, and this finding is then extrapolated to estimate the proportion of high to low frequency L2 words in the test-taker's lexicon. Laufer and Goldstein (2004) propose that the Lexical Frequency Profile in fact constitutes an indirect test of form-meaning association because only words that are used semantically correctly by the examinee are included in the profile generated by the test.

The Lex-30 test (Meara & Fitzpatrick, 2000) is a test of breadth of productive vocabulary knowledge that operates similarly to a word association task in that respondents are presented with 30 prompt words in response to which they must write four words they feel

are associated. The responses to the Lex-30 test are lemmatised and then scored based on their frequency (higher points for lower frequency words). One of the criticisms levelled at word association tasks is that the prompts themselves may influence the responses they beget (Moreno Espinosa, 2009). Moreover, it has been argued (Walters, 2012) that the Lex-30 only addresses one category of productive vocabulary knowledge as set forward by Nation's (2001) taxonomy (that of recall) whereas Fitzpatrick and Meara (2004) concede that the productive version of the Vocabulary Levels Test addresses five of these categories (meaning, appropriateness, collocations, grammatical position and orthographic form).

The Vocabulary Size Test (Nation & Beglar, 2007) measures breadth of receptive vocabulary knowledge; specifically, recognition of orthographic form, form meaning connection and, to a lesser degree, concept knowledge. The test, which uses a four-item multiple-choice format, can be used with L1 and L2 speakers and measures up to the fourteenth 1000 word level. Similarly to the levels test, the vocabulary size test derives the language it tests from native-speaker corpora; in this case, the BNC. The original test was monolingual, but a number of derived bilingual versions have been developed (e.g. Nguyen and Nation's English-Vietnamese version in 2011). Although Beglar (2010) went to great lengths to validate this test using Rasch analysis, the pervading criticism of the monolingual version of the Vocabulary Size Test is that, due to the long and grammatically complex multiple-choice definitions, it is not only a test of vocabulary knowledge but also of grammatical knowledge and reading skill (Nguyen & Nation, 2011).

2.2.5.2 Testing vocabulary depth

According to Read (2007), no test of depth of vocabulary knowledge has achieved the extent of acceptance that the Vocabulary Levels Test has for size. Read (1998) states that there are two approaches to testing depth of vocabulary knowledge: a developmental

approach, and a dimensional approach. The developmental approach pertains to the incremental nature of lexical acquisition and sets out to measure where on the developmental scale a learner's present knowledge lies. The dimensional approach refers to the many dimensions of word knowledge that have been posited by scholars (c.f. 2.1.2) and seeks to determine which dimensions a learner has mastery over. Clearly, to measure the spectrum of dimensions of word knowledge in a test would be severely to limit the number of lexical items under investigation, and as such it has been argued (Schmitt, 1998) that the dimensional approach is better suited to a research context than to use as a diagnostic tool.

The Computer Adaptive Test of Size and Strength (Laufer & Goldstein, 2004) aims to test both size and depth of vocabulary knowledge, but the test makers used the term 'strength' instead of depth and operationalised this construct at four hierarchical levels, from easiest to most difficult. The first level is passive recognition, whereby the examinee is given an L2 word as a prompt and then has to choose the correct L2 definition from four options. The second level is active recognition, for which the test-taker is given a L1 definition as a prompt and matches it with one of four L2 words. The third level is passive recall, whereby the respondent is provided with a L2 word as a prompt and must in response supply the L1 translation. Finally, the fourth level is active recall, for which the test-taker is prompted with a L1 word and must supply the L2 translation (the first letter of the L2 translation is provided).

The Vocabulary Knowledge Scale (Wesche & Paribakht, 1996) is a test of depth of vocabulary knowledge, which is derived from the developmental approach and utilises a 5-point self-report scale. The test was based on the principal that vocabulary knowledge is incremental and it aimed to test the development of those increments that a learner is able to report on (Pignot-Shahov, 2012). Laufer and Goldstein (2004) argue that the Vocabulary

Knowledge scale is in fact a test of form-meaning recognition (albeit indirectly) because four of the five levels on the self-report scale pertain to comprehension of meaning e.g. 'I have seen the word but I don't know what it means'. Moreover, Read (1998) has critiqued the 5-point scale in light of his doubt that the five levels are accurate and comprehensive representations of the stages of lexical acquisition. Qian's (1999, 2002, 2004) Depth of Vocabulary Knowledge measure is another vocabulary test based on the dimensional approach. The instrument measures three dimensions in terms of receptive vocabulary knowledge; synonymy, polysemy and collocation. Qian found scores on this measure to be more strongly predictive of reading comprehension than were vocabulary breadth scores, yet he himself states that the measure is only a partial operationalisation of depth of lexical knowledge. Another dimensional test of depth is the Word Associates Test (Read, 1993, 1998), which comprises a series of target words, each accompanied by six or eight different words, only half of which are associates (semantic or collocational) of the target word. As stated by Schmitt (2000), the Word Associates Test is one of the earliest tests of knowledge of multi-word lexical items.

2.2.5.3 Summary of testing vocabulary knowledge

To summarise, whilst a wide variety of different tests of vocabulary knowledge have been developed, there is as yet no 'gold standard' test for either size or depth. Monolingual vocabulary size tests of receptive knowledge are criticised because correct answers may be guessed at, yet few bilingual versions of the most standardised tests exist, and for research on learners of mixed L1s there is no option but to use a L2-L2 format. Furthermore, tests of vocabulary size have been criticised for measuring only certain facets of the construct of vocabulary knowledge, and also for measuring other constructs such as reading and grammatical skill. Similarly for tests of depth, a recurring theme is that the tests currently available measure some but not all of the dimensions or developmental stages of L2

vocabulary acquisition. Finally, as yet there are very few tests of multi-word lexical knowledge.

2.2.6 VOCABULARY KNOWLEDGE: EMPIRICAL EVIDENCE

In the following sections is detailed and discussed the empirical evidence with reference to the nature of vocabulary knowledge in two areas: the relationship between receptive and productive vocabulary knowledge, and the relationship between size and depth of vocabulary knowledge.

2.2.6.1 The relationship between receptive and productive vocabulary knowledge

Laufer (1988) conducted a cross-sectional investigation of the receptive and productive vocabulary development of n=48 L1 Hebrew secondary school learners of English as a L2, n=26 of whom had been studying English for six years and n=22 of whom had been studying English for seven years. She measured receptive knowledge using Nation's (2001) Vocabulary Levels Test, controlled productive knowledge using the productive version of the levels test (Laufer & Nation, 1999); and (3) free productive knowledge, using the Lexical Frequency Profile (Laufer & Nation, 1995). The more advanced students had a larger receptive vocabulary than the less advanced participants. There was a significant positive correlation between size of receptive knowledge and size of controlled productive knowledge [r=.78, *p*=.002], suggesting that the more words a learner knows receptively, the more they will know in controlled active terms. However, there was no statistically significant difference between the free productive knowledge of the two groups and no correlation found between receptive and free productive knowledge. Laufer (1998) proffered two explanations for the lack of significant findings with regard to free productive knowledge: (1) that the formal English instruction the children had received had not provided practice of using new lexis in free production, and (2) that the one year

difference between the groups was insufficient to detect differences in their free productive knowledge.

A decade later, Laufer and Paribakht (1998) studied the relationship between types of lexical knowledge and context of learning by testing, similarly to Laufer (1998), vocabulary knowledge in three dimensions: (1) receptive knowledge, measured using Nation's (2001) Vocabulary Levels Test; (2) controlled productive knowledge, measured using the productive version of the levels test (Laufer & Nation, 1999); and (3) free productive knowledge, for which the Lexical Frequency Profile (Laufer & Nation, 1995) was employed. The findings suggested that the order of developmental rate was receptive as the fastest, followed by controlled active, followed by free active. In terms of context, there was a significantly larger gap between the receptive and productive vocabularies of learners who had studied abroad than between those who had studied in the EFL setting: the EFL learners outperformed the study abroad learners in terms of both dimensions of productive knowledge, and the SAE learners demonstrated a richer receptive dimension of knowledge.

Another ten years on, Webb (2008) sampled n=83 L1 Japanese learners of L2 English, n=48 of whom had higher proficiency and n=37 of whom had lower proficiency, to investigate the relationship between receptive and productive vocabulary size. He administered two self-developed tests, which each tested vocabulary size at three frequency levels with lexical items selected from the COBUILD dictionary. The receptive test required respondents to write the L1 translation of the L2 target words. The productive test required respondents to write the L2 forms of words from a stimulus of their L1 meanings. As such, the measures tested knowledge of form and knowledge of meaning. ANOVA revealed that the receptive vocabulary size of the sample was significantly larger than the productive size [$F(1, 82)=194.67$, $p<.001$] overall, and in each of the three

frequency bands. Furthermore, the size of the gap between the two types of knowledge increased by frequency level, and for participants who had a smaller receptive size, the gap between their receptive and productive knowledge became increasing larger by frequency level than for learners whose receptive size was larger. As such, this study provides convincing evidence of a relationship between receptive and productive lexical knowledge: more receptively known words indicate more productively known words.

Thus, there is evidence to suggest that receptive vocabulary size is generally larger than productive size, and that the gap between these two dimensions of lexical knowledge is impacted by instruction and study. Moreover, the gap between receptive knowledge and controlled productive knowledge is smaller than the gap between receptive and free productive knowledge. Furthermore, there is evidence to suggest that there is relationship between size of receptive knowledge and size of productive knowledge. Additionally, and as stated by Laufer and Goldstein (2004), productive lexical knowledge appears to be more difficult to acquire than receptive lexical knowledge.

2.2.6.2 The relationship between size and depth of vocabulary knowledge

Qian (1999) sampled n=217 mixed L1 learners of English at the University of Toronto to investigate the roles of size and depth of lexical knowledge in reading comprehension. He measured size using the levels test (Nation, 2001) and depth using the Word Associates Test (Read, 1993, 1998) to find a significant correlation between scores on the two measures [r=.70, p<.01]. Two other studies have found significant positive correlations between measures of vocabulary size and depth as measured by a word association task: Nurweni and Read (1999) with n=350 L1 Bahasa Indonesia university level learners of L2 English, and Vermeer (2001) with n=50 Dutch kindergarten pupils.

Meara and Wolter (2004) sampled n=147 L1 Japanese learners of L2 English and administered a test of depth called V_Links, which had been developed by the researchers not only to test depth of knowledge of individual words but also the networks between them, in line with Read's (2004) construal of depth of vocabulary knowledge. They found a small correlation between depth (or organisation) and overall vocabulary size [$r<0.3$] but did not report whether this correlation reached statistical significance. Certainly, Wolter's (2005) further use of the V_Links measure failed to reveal any significant relationship to lexical breadth.

Gyllstad (2007) developed two receptive tests of collocations, COLLEX and COLLMATCH, to try to determine whether collocational knowledge is more closely related to depth or breadth of vocabulary knowledge. Sampling n=24 L1 Swedish learners of L2 English, he administered five measures of vocabulary knowledge, including the Vocabulary Levels Test (Nation, 1990) and Read's (1993, 1998) Word Association Test, to find that his collocation tests correlated with measures of both size and depth, but slightly more strongly with measures of size. However, there was very little difference between the relationship of collocation knowledge and breadth as with depth. Furthermore, the measures of breadth and depth were also highly correlated with one another. Gyllstad (2007) suggests that these relationships are indicative of depth and breadth, and collocational knowledge, all being facets of the same construct; that is to say, that the measures he employed in his study tested different aspects of the same construct. A replication of Gyllstad's (2007) study with a larger sample and a multiple regression analysis approach may shed further light on the relative contributions of different facets of vocabulary knowledge.

2.2.7 SUMMARY OF VOCABULARY KNOWLEDGE

As we have seen, vocabulary knowledge frameworks comprise reference to vocabulary size, depth of word knowledge and lexical access. Whilst implicit vocabulary acquisition is possible, there is not yet convincing evidence of the durability of implicitly learned words, and as such the argument for explicit approaches to vocabulary acquisition, in tandem with implicit learning from reading and listening, is strong. Explicit vocabulary instruction has been shown to lead to explicit lexical knowledge, but scholars have yet to determine the amount of explicit instruction necessary from which implicit knowledge can develop, and nor has the relationship between explicit and implicit knowledge clearly been defined. A range of size and breadth tests have been developed and used yet none are without critique and very few can measure knowledge of multi-word lexical items. In terms of breadth of vocabulary knowledge, receptive size has been shown to be larger than productive, and productive more difficult to acquire. Moreover, there is evidence to suggest that as receptive knowledge develops, so too does productive knowledge. With regard to size and depth of vocabulary knowledge, the literature points to a significant relationship between the two, and to a significant relationship between these dimensions and collocational knowledge.

2.3 STRATEGIC BEHAVIOUR

In this section the following areas of second language strategic behaviour are explored: the many definitions and operationalisations of the term ‘strategy’ in SLA scholarship; taxonomies of language learning strategies and vocabulary-related strategies; the relationships between strategic behaviour and individual and group differences; research methods and vocabulary-related strategic behaviour; and the current empirical evidence on the impact of vocabulary-related strategic behaviour on vocabulary gain.

2.3.1 DEFINING THE CONSTRUCT

Research into language learner strategies began apace in the 1970s as the focus in research moved away from aptitude and pedagogical approaches to language learning to hone in on how individuals can take control of their acquisition and use of a second language (Schmitt, 1997). Initially, strategies research concentrated on trying to establish the behaviours manifested by successful L2 learners (e.g. Stern, 1975; Rubin, 1975; Wong Fillmore, 1979) but as strategies were conceptualised further as goal-oriented, learner managed behaviour, so they became value neutral and the focus of the research community turned to defining and classifying strategic behaviour and to determining how strategies are effectively employed (Grenfell & Macaro, 2007). The attempts to define and conceptualise language learner strategies are hereby discussed with reference first to individual strategies and then to combinations of strategies.

2.3.1.1 Individual strategies

Although researchers almost unanimously agree that strategic behaviour plays a fundamental role in learning a second or additional language (e.g. O'Malley & Chamot, 1993; Ehrman & Oxford, 1990; Oxford, 1993), no one accepted definition of the term 'language learning strategy' exists. A great number of scholars have posited definitions over a period of five decades (Rubin, 1975; Rigney, 1978; O'Malley, Chamot, Stewner-Manzanares, Küpper & Russo, 1985; Ellis, 1985; Oxford, 1990; Griffiths, 2008, 2013) and it could be argued that the sheer number of the definitions and alternative terms proffered has served further to cloud the issue of what a language learning strategy actually is. Rubin's (1987:22) definition pertained to any actions that "contribute to the development of the systems which the learner constructs and affects learning directly". O'Malley and Chamot (1990:1) defined language learning strategies as "the special thoughts and

behaviours that individuals use to help them comprehend, learn, or retrieve...information", suggesting solely cognitive behaviour. Oxford's (1990:8) definition was "specific actions taken by the learners to make learning easier, faster, more enjoyable, more self-directed, more effective and more transferable to new situations", which suggests that she encompasses both physical and mental actions in her definition. Certainly, it has been argued that using observable action to define strategic behaviour is an invalid approach because one physical action could be the manifestation of any number of cognitive intentions (Sparks & Ganschow, 1993; Stevick, 1990).

Further disparity can be noted among the numerous definitions of 'strategy' with reference to Stevick's (1990) *Size Abstractness Dilemma*, in which it was argued that strategies can be very different in terms of the size of action (regardless of whether that action is cognitive or physical) and varied in terms of their abstractness (e.g. look up the meaning in a bilingual dictionary versus imagine a pictorial representation of the meaning). In response to the opacity of the construct, Dörnyei and Skehan (2003) proposed that in SLA research strategies should be defined as either neurological *or* cognitive *or* behavioural. In response, Dörnyei (2005) suggested that strategic behaviour be reconceptualised as self-regulatory goal-oriented behaviour. Macaro (2006) then proposed a cognitive framework for determining strategic behaviour which took account of Dörnyei's (2005) suggestion and which comprised four features: for Macaro (2006) strategies are (1) conscious and goal oriented; (2) located in working memory; (3) discrete (i.e. they do not involve the use of other strategies); and (4) situation and task specific yet transferable to other situations and tasks.

Takeuchi, Ikeda and Mizumoto (2012) set out to validate Macaro's (2006) framework using neuroimaging to detect activity in the prefrontal cortex, the hypothesised home of working memory (Baddelely, 1997). N=12 L1 Japanese learners of L2 English were given

three reading passages: for the first passage they were asked simply to read (control), for the second passage they were asked to scan for information and answer three True/False questions (experimental 1), and for the third passage they were asked to locate and underline topic sentences and then write a summary of the passage in the L1 or the L2 (experimental 2). Each participant completed each condition in both the L1 and the L2; that is, they read three L1 texts followed by three L2 texts, or *vice versa*. The findings revealed significantly more prefrontal cortex activity during the experimental conditions than the control condition, which provides evidence in support of Macaro's (2006) claim that cognitive strategy use is localised in working memory. Moreover, there was less neural activity whilst L1 strategies were deployed than when L2 strategies were in use, which further attests to Macaro's (2006) model.

In spite of the encouraging outcomes of Takeuchi, Ikeda and Mizumoto (2012) with regard to firm empirical evidence of the nature of strategies, limitations to the study prevent broad generalisations about its findings from being made. Primarily, the study only involved experiments with two distinct strategies, both of which derive from the 'cognitive' categorisation, and it may not be the case that other cognitive strategies or other categorisations of strategies, (e.g. socio-affective strategies) are similarly located. Furthermore, it may have been the case that the participants were manifesting strategies other than those under explicit investigation, even if (as posited by Macaro's model) all strategic behaviour is conscious mental activity. Finally, and as stated by the researchers themselves, there may have been an impact of text on the neural activity uncovered, particularly in light of the differing topics and difficulty levels of the six reading passages selected.

Whilst the disparity in the SLA research community regarding the nature of strategic behaviour is by no means resolved, a dearth in more recent years of newly posited

definitions for the term would suggest that Macaro's (2006) framework has gone some way toward encouraging a consensus on how strategy based scholarship might usefully conceptualise strategies going forward. However, as we shall see, one could certainly argue that the move from theory to practice (i.e. how strategies have been operationalised in research since 2006) has been less quick to respond.

As noted by Grenfell and Harris (2013), over the past decade research has begun to explore strategic behaviour from a sociocultural perspective; that is to say, language learner strategies have been defined in light of the social, cultural, political and economic contexts in which they are manifested. Oxford and Schramm (2007) argue that the ontological, epistemological and methodological differences between the psychological and sociocultural perspectives have the potential to muddy the waters further in terms of defining strategic behaviour, yet they also believe that the two approaches can be compatible because they share certain elements of interest. One mutual element of interest discussed in detail by Takeuchi, Griffiths and Coyle (2007) is that of context, or learning situation. Takeuchi *et al*. (2007) call for the inclusion of both cognitive and situational variables in research on strategic behaviour, and others (e.g. Yang & Jiménez, 2011) believe that strategic behaviour cannot be understood without consideration of context or social hierarchy.

2.3.1.2 Combinations of strategies

Much of the research into strategic behaviour in second language acquisition has focused solely on the use of individual strategies, yet scholars have argued that successful learners deploy combinations of strategies (Anderson, 2008; Griffiths, 2013; Grenfell & Harris, 2013) and that effective deployment of combinations of strategies is context and task-specific (Macaro, 2001, 2003). Strategic combinations can be dichotomised as either

clusters or *chains*, although, adding to the "definitional fuzziness" (Tseng, Dörnyei & Schmitt, 2006:79) surrounding language learner strategies, a number of scholars to date have tended not to differentiate between the two (e.g. Hurd & Lewis, 2008; Oxford, 2011).

Clusters refer to the use of multiple strategies in no predetermined order. That is to say, the learner has not based on previous experience or hypotheses about strategic cooperation the strategies they deploy in a cluster. Rather, a cluster of strategies is representative of a learner applying a variety of different strategies in order to see which helps them to achieve their task and/or situation specific goals. Chains, on the other hand, refer to systematic sequences of strategies; that is, combinations of strategies that the learner has tried in a set order previously and that exist in their repertoire for deployment in particular tasks and scenarios. One could argue that, similarly to the singular nature of multi-word lexical items in the lexicon, strategic chains are stored and retrieved by the learner as a 'chunk'. In other words, at the point of deployment the learner is not consciously aware of the individual strategies that comprise the chain.

2.3.2 TAXONOMIES OF LANGUAGE LEARNER STRATEGIES

A number of taxonomies of language learner strategies have been posited. One early strategies taxonomy was that of O'Malley and Chamot (1990). They arrived at a tripartite classification structure for language leaning strategies that encompassed metacognitive (defined as involving overviewing, planning and regulating learning), cognitive (task-specific manipulation of information to remember or acquire it) and socio-affective strategies (involving interpersonal relationships and the regulation of one's emotion). Chamot, Kupper and Impink-Hernandez (1987), who designed a Language Learning Strategies Inventory comprising 48 individual strategies, which they used to elicit the strategic behaviour of L2 learners of Spanish and Russian, posited an almost identical

taxomomy. They classified their strategies into three categories: metacognitive (defined as involving self-regulation in terms of planning, monitoring and evaluating learning), cognitive (pertaining to what learners do to achieve a goal in a particular task) and social and affective (comprising interaction and emotional control). These early taxonomies suffer from indistinct classification boundaries in that, for example, if engaged in an interactive task (such as describe and draw), then the cognitive strategies as defined by these taxonomies are also socio-affective strategies.

Rubin's (1987) taxonomy of language learner strategies is another three-part classification system, comprising learning strategies (cognitive and metacognitive), communication strategies (to repair breakdowns in interaction), and social strategies (providing exposure to and practice of the L2). Once more there is overlap between the categories, as communication and social strategies could be one and the same. Stern's (1992) taxonomy was divided into five categories: management and planning strategies (regarding the learner's self-directed learning intentions), cognitive strategies (related to manipulation of learning materials), communication-experiential strategies (to maintain conversational flow), interpersonal strategies (interaction and 'cooperation' with native speakers), and affective strategies (promotion of positive emotion and management of negative emotion). Here again the classifications are not distinct e.g. interpersonal strategies may well encompass communication-experiential strategies.

The most influential of the taxonomies of language learning strategies has been without a doubt that of Oxford (1990), from which was formed the *Strategy Inventory for Language Learning* (SILL), a questionnaire tool that measures the frequency with which language learners engage in different types of strategic behaviour. The SILL was originally designed for use with students at the Defense Language Institute in California but has pervaded strategies research worldwide to that point that it had been used in well over fifty different

studies (Oxford & Burry-Stock, 1995) and has been translated into over twenty languages (Oxford, 2011).

The SILL requires respondents to state on a five-point scale the frequency with which they engage in the 62 strategies it describes. The strategies in the SILL are subdivided into six categories (cognitive, metacognitive, memory, affective, compensation and social) and into two classifications (L2 only or L1 supported). Memory strategies are defined as pertaining to the memorisation of lexical items, cognitive strategies are defined as referring to text comprehension and production, compensation strategies involve the learner trying to overcome problems caused by gaps in their knowledge, metacognitive strategies oversee language learning processes, affective strategies manage the emotions of the learner, and social strategies are interactive in nature. Unfortunately, as with the taxonomies of O'Malley and Chamot (1990), Chamot *et al*. (1987), Rubin (1987) and Stern (1992), Oxford's classification structure is not entirely rigorous because there is overlap between her categories (e.g. social and cognitive). Furthermore, and as stated by Dörnyei (2003), the frequency of a strategic action may be wholly unrelated to its efficacy: that is to say, a learner may engage frequently in a strategic behaviour which has little to no positive impact on his or her second language acquisition.

Ehrman and Leaver (2003) categorised strategic behaviour based on the level of processing the strategies involve. Their three categories were: surface strategies (which are task-specific and do not require deep cognitive processing or emotional load), achievement strategies (which serve to build interpersonal relationships), and deep strategies (which involve making links between novel and existing input in order to commit new linguistic knowledge to long term memory). The classification labels in this model might sound different, but one could definitely argue that they conform to the 'cognitive-metacognitive-socio-affective' basis of all of the pre-SILL taxonomies. For example, achievement

strategies in this model are defined as interpersonal and therefore they are by their very nature social strategies.

In summary, the similarities between the language learner strategy taxonomies that have been posited are that they all include classifications of cognitive, metacognitive and socio-affective strategies, but that they none of them make use of stringent classification boundaries. From the basis of some of these taxonomies were developed taxonomies specific to vocabulary-related strategic behaviour, which are discussed in detail in 2.3.4.1.

2.3.3 STRATEGIC BEHAVIOUR AND INDIVIDUAL AND GROUP DIFFERENCES

The individual and group differences that have been investigated with reference to strategic behaviour include age, gender, culture, and language learning experience/proficiency level. The evidence amassed with regard to each of these variables is now discussed and evaluated.

2.3.3.1 Age

Schmitt's (1997) study of n=600 L1 Japanese leaners of L2 English divided the sample into four different groups based on age: junior high school pupils, senior high school pupils, university students, and adults. He used his taxonomy of vocabulary strategies (detailed in 2.3.4.1) to devise a survey which he administered to n=600 L1 Japanese leaners of English. The survey was translated into Japanese and the respondents were required to state 'Yes' if they used each strategy or 'No' if they did not. They also stated whether or not each strategy was helpful, or whether, in the case of strategies they reported not using, they thought the strategy would be helpful if used. The findings revealed that the older the participant, the more they engaged in meaning-based strategic behaviour than form-related memorisation approaches.

In a recent factor analysis approach to strategies research, Tragant, Thompson and Victori (2013) sampled n=1,975 L1 Catalan pupils of L2 English, n=550 of whom were middle grade pupils (12-14 years old) and n=1,425 of whom were upper grade learners (15-17 years old), in response to Grenfell and Macaro's (2007) call for research attention on this population of L2 learners. The study sought to validate a language learning strategies questionnaire developed by Tragant and Victori (2012), which had been designed for employment in the EFL context. The questionnaire comprised 73 items and utilised a six-point frequency scale from 'never/hardly ever' to 'always/almost always'.

Exploratory factor analysis extracted a two-factor model: the first factor comprised strategies which involved "deep processing" (2013:102) in reading, writing or listening, and the second factor comprised strategies that pertain to the memorisation of grammar and lexis. This model was used to determine whether there were differences between the two sample groupings. The questionnaire data was then analysed using confirmatory factor analysis to reveal that there was a significant difference between the groups on the deep-processing skills-based strategies [$\Delta\chi^2(1)=7.223$, $p<.007$], and on the memorisation strategies [$\Delta\chi^2(1)=76.350$, $p<.001$]: similarly to the findings of Schmitt (1997), the upper grade (older) learners reported more frequent use of the deep processing skills-based strategies and less frequent use of the memorisation strategies than did their lower grade (younger) counterparts.

Thus, the evidence on age and vocabulary-related strategic behaviour appears to suggest that older learners focus more on meaning-derived strategic approaches whereas their younger counterparts are more form-focused, and particularly with regard to memorisation.

2.3.3.2 Gender

The research on the relationship between gender and language learning strategy use suggests a strong link between the two and has tended to find that female learners report more frequent strategy use (e.g. Ehrman & Oxford, 1988; Bacon & Finnemann, 1992; Oxford & Niykos, 1989; Sheorey, 1999) and use of different strategies from male learners (e.g. Bacon, 1992; Bügel & Buunk, 1996; Gu, 2002; Khalil, 2005; Liyanage & Bartlett, 2012). For example, Ehrmann and Oxford (1988) administered Oxford's (1990) SILL to reveal that female learners reported using four classifications of strategies more often than males: self-management strategies, authentic language use strategies, general study strategies and strategies involving searching for and communicating meaning. In another study that employed the SILL, Goh and Kwah (1997) found that female learners of English in China reported using cognitive and affective strategies more frequently than did males. Khalil (2005) administered the SILL to n=378 Palestinian learners of L2 English to find that women reported more frequent use of socio-affective strategies than men. Females have also been found to be more open to using new vocabulary-related strategies than men (Oxford, Lavine, Felkins, Hollaway & Saleh, 1996; Young & Oxford, 1997; Gu, 2002).

Some research on gender and strategic behaviour has found no difference between male and female learners. For example, Park's (1997) study of East Asian high school learners of English failed to reveal any impact of gender, and Kaylani's (1996) investigation of the social strategic behaviour of n=255 Jordanian high school pupils suggested that social strategies were used equally as frequently by boys and girls. There is also research to suggest that males use translation strategies more frequently than women (e.g. Bacon, 1992; Liyanage, 2004). Additionally, neither Wharton's (2000) study of bilingual

Singaporean language learners, nor Shmais' (2003) study of Palestinian learners of L2 English found any effect of gender on strategic behaviour.

In more recent gender-oriented strategies research, Liyanage and Bartlett (2012) conducted an investigation into the impact of gender on the strategy use of n=866 learners of L2 English in Sri Lanka. The researchers adapted the *Language Learning Strategies Inventory*, a 48-item questionnaire designed by Chamot, Kupper and Impink-Hernandez (1987) to measure the frequency of use of 16 different language learning strategies in a mixed methods study of L2 learners of Russian and Spanish. However, in qualitative data collection by Chamot *et al*. (1987) the participants mentioned ten strategies that had not been included in the inventory. As such, Liyanage and Bartlett (2012) included these ten strategies to form a 63-item questionnaire of 26 different strategies, which were divided into three subcategories: metacognitive, cognitive and socio-affective. The instrument was translated into the L1s of the sample (Tamil and Sinhala) and the respondents were required to choose one option from a four-point frequency scale (from 'almost never' to 'almost always') in response to each item. Across all three categories females reported more frequent strategy use than males ($p<.001$). However, when the relationship between gender and the 26 individual strategies was conducted, there was only a significant difference between the frequency of use on 16 strategies, and in the case of seven of these 16 strategies it was men who reported more frequent use.

In conclusion, the literature on gender and language learning strategy use tends to suggest that females use more strategies than males, but there is evidence to suggest that this may not be the case for all age groups and learning contexts. Moreover, the empirical evidence has for the most part been gathered using instruments that use frequency scales and that force respondents to generalise their strategic behaviour across situations, and as such the reliability of the findings here discussed should be treated with caution.

2.3.3.3 Culture

Much of the literature on the relationship between culture and strategic behaviour has honed in on Asian learners of L2 English. Some research has suggested that Asian learners have a preference for using rote learning strategies. For example, O'Malley, Chamot, Stewner-Manzanares, Küpper and Russo (1985) found that Asian learners of L2 English were unwilling to use imagery and grouping strategies to help them learn unfamiliar vocabulary. Instead, the Asian learners manifested rote memorization strategies and were able to apply these strategies to gain significantly more lexis than the control group in the O'Malley *et al.*'s (1985) study, who had received explicit training in more complex strategic approaches. However, in a number of more recent studies that have used the SILL (Oxford, 1990), Chinese learners of English have been found to make frequent use of metacognitive, cognitive and compensation strategies more so than memorisation strategies (Goh & Kwah, 1997; Bremner, 1998; Peacock & Ho, 2003). Similarly, Grainger (1997) and Mori (2007) found that L1 Japanese learners of L2 English used social strategies most frequently, and affective and memorisation strategies least frequently. However, whether strategic behaviour specific to any particular culture is impacted directly by the culture itself or the educational norms in that culture is difficult to determine. For example, in a study of the strategic behaviour of Asian learners of English, Bremner (1998) suggested that his university level sample found it difficult to use social strategies due to the focus in their six years of prior formal English instruction at school on grammar, vocabulary and reading.

2.3.3.4 Language learning experience/proficiency level

The literature on proficiency level and strategic behaviour intimates a positive relationship between the two. For example, in an investigation of strategic behaviour in L2 reading,

Razi and Grenfell (2012) employed a think aloud protocol to study n=12 L1 Turkish secondary school pupils who were L2 learners of English. The researchers positioned their study with reference to the *threshold hypothesis*, which posits that learners require a certain level of linguistic knowledge (that is, the requisite lexical and syntactic knowledge base to facilitate bottom up processing) in order to be able successfully to deploy language learner strategies and transfer their L1 reading skills to the L2. The study aimed to reveal the relationship between low levels of linguistic knowledge and strategy deployment in L2 reading. The qualitative data suggested that the sample was constrained in terms of the success of their strategy deployment in reading by their lack of morphological and lexical linguistic knowledge, which was taken as evidence in support of the threshold hypothesis and thus suggests that lower proficiency learners are less adept at strategy use than higher proficiency learners. However, as only two extracts from one participant's think aloud were included in the article, the weight of data in support of this finding is unclear. Schouten-van Parreren (1989) found, in another study of vocabulary strategies in L2 reading, that lower proficiency learners failed to use textual context to help them infer the meaning of unfamiliar words, focusing instead on the unfamiliar word itself.

Studies have revealed that more highly proficient learners are wont to manifest a greater number of strategies than their less proficient counterparts (e.g. Khalil, 2005; Magogwe & Oliver, 2007; Park, 1997; Shmais, 2003). There is also evidence to suggest proficiency level plays a role in the types of strategies that L2 learners employ. For example, Fan (2003) sampled n=1,067 L1 Cantonese learners of L2 English in Hong Kong and divided her participants into three groups (low, middle and high) based on their scores on a version of Nation's (1990) receptive Vocabulary Levels Test. Participants in the high vocabulary knowledge group used certain strategies more frequently than did the other two groups. Her more highly proficient group employed a significantly greater number of sources to

determine meaning, guessed more often, used their dictionaries more, and used existing word knowledge more frequently. Fan (2003) took this finding to suggest that the types of strategies used by highly proficient learners should be taught explicitly to lower proficiency learners. Hong-Nam and Leavell (2006) administered the SILL to n=55 mixed L1 learners of L2 English at three proficiency levels and found that their advanced level participants reported using social strategies most frequently, whilst the beginner and intermediate participants used metacognitive strategies more frequently then strategies from any other category.

Researchers have also argued for a relationship between language learning experience and strategic behaviour. Oxford and Nyikos (1989), in their study of American university students, and Ramirez (1986) in a study of adolescent L1 English learners of L2 French found that the number of years a learner had studied English had a significant impact upon their strategic behaviour. Khalil's (2005) study of university and high school learners of English revealed the university students were using more language learning strategies than the high school pupils. Studies have also suggested that the type of language instruction a learner has experienced may impact on the strategic choices (e.g. Bremner, 1998).

To summarise, it appears from the literature that linguistic knowledge enables strategic behaviour and that there is a positive correlation between the two. However, no evidence in support of causation has been uncovered (except in intervention studies) so it is not clear whether strategic behaviour is the chicken or the egg with reference to proficiency level in naturalistic settings. Furthermore, whilst research has investigated both language learning experience and proficiency, there is no current evidence to distinguish between the relative influences on strategic behaviour of these two variables. As such, we cannot yet say which plays a more vital role in how learners try to control their L2 acquisition and use.

2.3.4 VOCABULARY-RELATED STRATEGIC BEHAVIOUR

Vocabulary-related strategic behaviour describes the application of language learning strategies in response to newly encountered L2 lexis. A distinction has been made (Schmitt, 1997, 2000) between vocabulary strategies that serve to aid comprehension of a lexical item (vocabulary comprehension or 'discovery' strategies) and strategies which serve to consolidate knowledge about new items (vocabulary acquisition strategies). The term 'vocabulary-related strategic behaviour' is used in the present study to encompass both comprehension and acquisition strategies. Vocabulary-related strategic behaviour is hereby discussed with reference to the classification systems posited by the research community, the research methods used in vocabulary strategy research, and the current evidence on the relationship between vocabulary-related strategic behaviour and vocabulary gain.

2.3.4.1 Taxonomies of vocabulary-related strategies

An early taxonomy of vocabulary-related strategies was posited by Williams (1985), with reference to the strategies available to second language learners who encounter unfamiliar lexis in reading. He identified five types of strategies that he believed could form the basis of explicit strategy instruction. The five strategies were: contextual inferencing, word analysis, searching for synonyms, unchaining nominal compounds (e.g. using the component parts of 'redhead' to determine its meaning) and detecting lexical familiarisation. Williams (1980) defines lexical familiarisation as the means by which an author intentionally familiarises the reader with the meaning of a word or phrase using verbal, illustrative or numerical devices. As stated by Nation (2001), this classification system is interesting in that some of the strategies it contains can be applied not only to newly encountered words but also to the reinterpretation of previously learned words.

Nation (2001) posited a theoretical taxonomy of vocabulary strategies that comprised eleven types of strategies divided into three strategy 'classes' as illustrated in table 2.2.

Table 2.2: Nation's (2001:218) taxonomy of vocabulary learning strategies

Strategy Class	Strategy Type
Planning: choosing what to focus on and when to focus on it	Choosing words
	Choosing the aspects of word knowledge
	Choosing strategies
	Planning repetition
Sources: finding information about words	Analysing the word
	Using context
	Consulting a reference source in L1 or L2
	Noticing parallels in L1 and L2
Processes: establishing knowledge	Noticing
	Retrieving
	Generating

One could argue that Nation's (2001) taxonomy requires further subdivision in order to clarify the intention of the learner (their strategic 'goal') as they deploy any one type of strategy. For example, 'Consulting a reference source in L1 or L2' fails to specify the kind information the learner hopes to find, or why. Are they looking for a L1 translation? Are they looking for an example of the word in context? Do they want to know about one particular meaning of the word, or are they hoping to investigate its polysemy? Furthermore, more specificity is required in terms of the reference source itself: is it a dictionary, a person, a search engine? Does it provide only L1 or L2 information about the word, or both? Questions such as these arise from the majority of Nation's (2001) taxonomy so, whilst it is a useful conceptualisation of vocabulary-related strategies, it is by no means a global and comprehensive representation of the term.

Gu and Johnson (1996) proposed a taxonomy of vocabulary learning strategies that comprised 91 strategies subdivided into two parent categories (cognitive strategies and metacognitive regulation strategies) and six subcategories (vocabulary learning beliefs, guessing, memorisation, dictionary, note taking and activation). Schmitt's (1997, 2000)

taxonomy of vocabulary-related strategies was based on Oxford's (1990) and comprised 58 strategies that he subdivided into five categories: (1) determination, pertaining to a learner working out by themselves the meaning of a word/phrase; (2) social, whereby the learner interacts with others to determine the meaning of a word/phrase; (3) memory, describing actions taken by the learner to move the word/phrase into the long-term memory; (4) cognitive, which relates to the learner working with the word in some way; and (5) metacognitive, which refers to a learner planning and evaluating their vocabulary learning). Schmitt's (1997) taxonomy-based survey, which he administered to n=600 L1 Japanese leaners of English, was translated into Japanese required respondents to state 'Yes' or 'No' depending on whether or not they used each strategy. They also stated whether or not they thought each strategy was, or would be, helpful. The results revealed that the strategies that most participants reported using were using a bilingual dictionary and asking classmates to determine the meaning of a newly encountered word.

In terms of the categories into which Schmitt (1997) placed his 58 vocabulary-related strategies, there is no evidence that the strategies in each have the same underlying features because no factor analysis was performed on the survey data. Furthermore, and not dissimilarly to Oxford's (1990) taxonomy, there is overlap between some of the category boundaries. For example, Schmitt defines his cognitive category as comprising strategies that are similar to memorisation strategies but that involve less "manipulative mental processing" (1997:14) and instead focus more on repetitive and mechanical approaches to studying vocabulary, yet he includes in his memorisation category the strategy of committing to memory the individual words that make up a multi-word unit, which one could argue is itself a mechanical approach to the acquisition of a lexical chunk.

2.3.4.2 Research methods and vocabulary-related strategic behaviour

McDonough (1995) proposed that methodological approaches to researching vocabulary-related strategic behaviour can usefully be dichotomised as either *indirect* or *direct*. Direct methods pertain to research participants providing a self-report, either concurrently or retrospectively, of their strategic behaviour during a task, whereas indirect methods involve research participants stating the extent of their identification/agreement with researcher-devised statements.

2.3.4.2a *Direct research methods*

Paribakht and Wesche (1999) conducted a study of the strategies used by n=10 university level L2 learners of English when they encounter unfamiliar vocabulary during reading tasks. They took a tandem methodological approach in that they used think-aloud protocol, whereby the participant verbalises their thoughts whilst engaged in a task, in conjunction with stimulated recall, whereby participants are prompted to relive an event and report the mental processes they experienced (Gass & Mackey, 2000). The participants comprised a variety of L1s (including Farsi, French and Vietnamese) and were at intermediate proficiency level. "Several weeks before" (1999:201) the think aloud protocol, the sample had each been shown a reading passage about acid rain and were asked to indicate which words in the text were unfamiliar to them. This constituted the pre-test. The participants received training from the researchers with regard to the processes of think aloud and were then asked to read the target text (which appears from their reporting to be the same text that was used in the pre-test) and to complete two tasks. The first task was to answer comprehension questions about the reading passage, and the second task was verbally to summarise each paragraph whilst reading the passage. During the tasks the sample had access to a monolingual English dictionary.

Delayed stimulated recall protocol was also employed in Paribakht and Wesche's (1999) study in that the researchers asked their sample after the tasks had been finished to answer questions about the difficulty level of the reading, their interest in it, and how easy it was to work with the words they had chosen to strategically engage with. The data revealed that the sample chose not to engage with the majority (56%) of the words they did not know. Three vocabulary-related strategies were reported when learners did choose to engage with an unknown word: retrieval (i.e. repeating the word aloud or rereading it to try to encourage any possible recall of meaning), appeal for assistance (i.e. asking the researcher for the meaning) and inferencing (i.e. guessing from textual/background context). Of these three, lexical inferencing was by far the most commonly reported strategy, comprising almost 80% of all strategic behaviour reported. One possible improvement to the methodological approach taken in this study would be to have participants think aloud in the L1 so that their verbalisations are unmediated by their capacity to express themselves precisely in the L2. However, with a sample of mixed L1s this might rarely be possible.

Sampling n=93 North American university level learners of L2 Spanish, Barcroft (2009) conducted a study into the vocabulary-related strategies used by his sample in intentional (i.e. explicit) vocabulary learning. His research methodology was direct in that it involved self-reporting in retrospect (i.e. stimulated recall). The sample were first asked to "do their best to learn" (2009:79) a series of twenty-four word-picture pairs. Then they were tested on their knowledge of those pairs in two formats: in the first they were shown the picture and asked to recall the L2 word, and in the second they were shown the L2 word and were asked to supply an L1 translation. After completing the tasks the sample were asked to report on the strategies they had used to try to learn the new words. Unfortunately, Barcroft (2009) fails to state the time period that elapsed between the end of the testing

and the start of the self-reporting session and therefore the extent to which his participants' self-reports may have been impacted by memory is unclear.

Briggs (2014), in her study of the vocabulary-related strategies used by learners of English in informal L2 contact, sampled n=95 mixed L1 participants who were on a SAE in Oxford. She noted that scholars, such as Tseng, Dörnyei and Schmitt (2006), had stated that the potential of stimulated recall in terms of investigating strategic behaviour had not yet fully been exploited, and that simulation as opposed to stimulation may overcome the methodological weakness inherent in retrospective reporting (Gass, 2001): that is, that in stimulated recall, such as for Barcroft (2009), the reliability of the self report is influenced by the temporal proximity of the recall to the task. In light of these statements, Briggs (2014) devised the Opportunities With Language Simulator (OWLS), a research tool that comprised computer-based simulations of the five most identified with informal L2 contact scenarios as defined by her sample. Each simulation contained a target word in response to which she hypothesised the sample would deploy strategies. She used the tool as a stimulus in semi-structured interviews to elicit the task and situation-specific behaviour that a subsample of n=23 participants were manifesting.

The data gathered by Briggs (2014) revealed that strategic behaviour in informal L2 contact is subject to the influence of a vast array of factors, and the interplay between them: the contextual intention of the learner (i.e. the overarching reason they are engaging in a particular scenario); the physical and/or social constraints of the scenario (e.g. participants did not report using dictionary-based strategies in the cinema), their strategic intention (e.g. do they want simply to comprehend the lexical item or do they want to try to remember/use it?), and even the target words themselves played a role in determining the strategic action of the sample. Although this methodological approach highlighted the complexity of the nature of strategic behaviour in a way that other methods have as yet

failed to, Briggs (2014) failed to include in her instrument any multi-word target lexical items, and as such her operationalisation of vocabulary was incomprehensive. Moreover, and similarly to Paribakht and Wesche (1999), Briggs' (2014) participants were thinking aloud in the L2 and as such the sophistication of their verbalisations may have been limited.

2.3.4.2b *Indirect research methods*

Indirect methods have more commonly been used to investigate vocabulary-related strategic behaviour than have direct methods, most likely because indirect methodologies lend themselves more easily to larger samples and are more straightforward in terms of data analysis. Gu and Johnson (1996) devised their Vocabulary Learning Questionnaire based on their taxonomy of vocabulary-related strategies and administered the research instrument to n=850 Chinese learners of L2 English to determine whether there was a relationship between strategy use as measured by the questionnaire and lexical range and language learner beliefs. Respondents were required to state on a seven-point scale how true of them each strategy was, and as such the extent of the sample's identification with their 91 strategies was measured.

Gu and Johnson (1996) administered the Vocabulary Learning Questionnaire to their sample in tandem with two vocabulary size tests, which included Nation's (1990) levels test and a baseline proficiency measure (a composite of scores from three different types of test). The findings revealed that meaning-oriented strategies were more highly identified with than rote learning strategies. Cluster analyses were used to arrive at a strategic profile of the sample to find that the participants identified more highly with meaning oriented strategies than with rote learning strategies. The researchers also employed cluster analysis to create a strategic profile for each participant, yet these profiles were not comprehensive

because the questionnaire the researchers had administered does not capture the use of strategic combinations (i.e. clusters and chains). Moreover, and similarly to Oxford's (1990) SILL and the questionnaire used by Fan (2003), their questionnaire forces respondents to generalise their actions across contexts (Winne & Perry, 2000; Tseng, Dörnyei & Schmitt, 2006): for example, it is conceivable that a learner may use dictionary-based strategies very frequently in the classroom setting, but they may not carry a dictionary with them outside of class.

In a more recent direct approach, Zhang and Li (2011) devised a 60-item vocabulary strategies questionnaire in which the strategies they included were divided into the same six categories as utilised by Oxford (1990): cognitive, memory, compensation, affective, social and metacognitive. The researchers do not state their operationalisation of the term 'vocabulary strategy', nor do they state from whence their 60 strategies derived. Their respondents (n=296 Chinese learners of L2 English) were required to choose one option from the following 5-point 'how true of me' scale for each item on the questionnaire (2011:144):

1. Never or almost never true of me
2. Generally not true of me
3. Somewhat true of me
4. Generally true of me
5. Always or almost always true of me

This scale was an improvement on the purely frequency based scales used by other strategy questionnaires, such as Oxford's (1990) SILL, in that points 2, 3 and 4 did not require the respondent to estimate the frequency of occurrence of each strategic behaviour. However, strangely, points 1 and 5 did require the respondent to state the frequency with

which they do something, and to generalise that action across tasks and situations. That is, the points in the scale are not equal to one another; they do not measure the same construct, and consequently the value of the data yielded from them is severely impacted.

The researchers used their questionnaire data to conduct Confirmatory Factor Analysis (CFA) with reference to four different factor models: (1) a six-factor model based on Oxford's (1990) six-category taxonomy; (2) a four-category model in which cognitive, compensation and memory strategies were combined to make one cognitive category, to be confirmed together with a metacognitive, social and affective category each; (3) a three-factor model based on O'Malley and Chamot's (1990) taxonomy, comprising social/affective, cognitive and metacognitive strategies; and (4) a two-factor structure of cognitive and metacognitive strategies similar to Gu and Johnson's (1996) model. Moreover, Zhang and Li (2011) conducted Exploratory Factor Analysis (EFA) on the questionnaire data to see what factor structure emerged. EFA revealed a six-factor structure; hardly surprising given that the questionnaire was written based on Oxford's (1990) six categories. The researchers did not report the findings of the CFA.

Unfortunately, many of the 'strategies' included in Zhang and Li's 60-item questionnaire are not in fact strategies; certainly not according to any of the definitions posited for the term in section 2.2.1. For example, the researchers stated that "I have a general idea of how many English words I know" (2011:147) is a metacognitive vocabulary strategy, but this statement neither describes a physical nor a cognitive action that pertains to either the comprehension or acquisition of a lexical item. Therefore, whilst a factor analysis approach to determining the underlying structures in vocabulary strategy data is certainly pertinent, first a stringent operationalisation of the key term should be stated and adhered to. Moreover, care should be taken not to impose any presupposed structure on the data.

2.3.4.3 Vocabulary-related strategic behaviour and vocabulary gain

A number of studies have explored the relationship between vocabulary-related strategic behaviour and vocabulary gain. Gu and Johnson (1996), in their administration of their Vocabulary Learning Questionnaire to n=850 Chinese learners of English in tandem with a proficiency measure that comprised a composite of scores from three different tests, found that two metacognitive strategies (self-initiation and selective attention) and one cognitive strategy (verbal repetition) were significant predictors of vocabulary breadth. This finding mirrors that of Wang, Thomas, Inzana and Primicerio (1993), who found that, given a list of words to learn, participants who employed more repetition strategies achieved higher scores on a recall test. However, interpretations of Gu and Johnson's (1996) findings should bear in mind that their measure of strategic behaviour was based on self-report, and that self-perception may well differ somewhat from reality.

Lawson and Hogben (1996) overcame the methodological weakness of self-report by using think aloud protocol to collect data on what their sample of n=15 Australian learners of Italian did in response to newly encountered lexis in explicit vocabulary learning, and then sought to determine whether there was a relationship between that behaviour and recall. They coded their think aloud data into four strategy classifications: repetition, word feature analysis, simple elaboration and complex elaboration. The vocabulary test involved productive recall of form-meaning connection in that respondents were shown 12 words in the L2 and were required to write the meaning of each. There was a strong, statistically significant relationship between frequency of strategy use and recall scores [r=0.83, p<.01]. In support of the findings of Gu and Johnson (1996) and Wang *et al*. (1993), Lawson and Hogben (1996) found that repetition strategies were positively related to higher scores on the recall test. Higher scores were also positively correlated to elaboration strategies such as paraphrasing [r=0.62, p<.05] and mnemonics [r=0.52, p<.05]. However,

the recall test in Lawson and Hogben's (1996) study was conducted only 30 seconds after the final vocabulary learning session and so, as no delayed recall test was administered, the durability of these findings is not established.

In an investigation of the vocabulary-related strategies used by learners of English in two distinct settings, Kojic-Sabo and Lightbown (1999) conducted a cross-sectional study of n=90 learners of L2 English to determine whether there was a relationship between vocabulary size and self-reported strategic 'approach' to L2 vocabulary learning. They administered a questionnaire based on the work of Sanoui (1992) to determine the sources of the participants' newly encountered vocabulary, the opportunities they had to practice these words, and their preferences with regard to working with newly encountered words. The questionnaire required participants to indicate their position on a cline (from self to teacher initiated) for each questionnaire item. The vocabulary measure employed was Meara and Buxton's (1987) Yes/No test, which contained 120 words of which 60 were peseudowords. The findings revealed a strong relationship between vocabulary size and learner independence (i.e. how far their strategic activity was self-initiated) and time spent in vocabulary-strategic activity, and therefore the findings suggest a positive relationship between vocabulary-related strategic behaviour and vocabulary gain. Certainly, Kojic-Sabo and Lightbown (1999) state that their findings support those of Sanoui (1992), who found that extensive use of strategies positively correlated with language learning success. However, the researchers' definition of the term 'strategy' in this study was no by means stringent. For example, they include as strategy items in their questionnaire activities such as "listening to the radio" and "watching English news" (1999:179), which one could argue refer more to informal L2 contact than to strategic behaviour. Moreover, the use of the Yes/No test as the measure of vocabulary size means that the scores on this test may have been inflated due to guessing. As such, Kojic-Sabo and Lightbown's (1999) study in

fact may reveal very little about the relationship between vocabulary-related strategic behaviour and vocabulary gain.

Taking a broader, more curriculum-based view of vocabulary-related strategies, Moir and Nation (2002) conducted a series of interviews with n=10 adult L2 learners of English on an intensive 12-week course in an ESL setting to determine what strategies they were using to help them comprehend and acquire L2 vocabulary, what factors were helping to determine their strategic choices, and to what extent those choices were effective. The sample was required to choose 30-40 L2 words to study for each week that they were on the 12-week course. The following week they were tested on their depth of knowledge of ten of those words chosen at random by the researchers. The tests were based upon Wesche and Paribakht's (1996) Vocabulary Knowledge Scale. Unfortunately, Moir and Nation (2002) found that on the whole the learners in their study did not use vocabulary-related strategies effectively and in fact forgot most of the words they had studied. The researchers believe that their findings resulted from the learners in their study working only to short term goals that had no personal significance, to a lack of understanding of the many dimensions of word knowledge, and to knowledge of only a limited range of vocabulary-related strategies.

Adams (2006) studied the impact of a two to four month SAE on the strategic behaviour of American learners of a variety of L2s in a range of host countries, such as Brazil, Austria and France. She administered the SILL (Oxford, 1990) and a self-report proficiency test pre and post the SAE. A positive correlation was found between self-reported gains in proficiency and self-reported gains in identification with strategies. However, as the SILL forces respondents to generalise their strategic action across situations (Winne & Perry, 2000; Tseng, Dörnyei & Schmitt, 2006) and proficiency was measured via self-report,

these findings do not provide particularly strong evidence to support a hypothesised relationship between the two.

2.3.5 SUMMARY OF STRATEGIC BEHAVIOUR

As we have seen, the literature on strategic behaviour has suffered from the lack of consensus regarding a singular conceptualisation and operationalisation of the term, yet taxonomies of language learner strategies and vocabulary-related strategies abound. The evidence to date indicates that linguistic knowledge enables strategic behaviour and that there is a positive correlation between the two. It appears from the literature that age may beget more meaning-focused strategic behaviour and females may use more strategies and more frequently than males. Some effect of culture has been suggested, but its influence separate from educational background has not been defined. Direct research methods such as think aloud, stimulated recall and simulation have garnered rich, detailed qualitative data about strategy use by language learners, whereas indirect methods have largely employed questionnaires, which are useful in drawing links between strategic behaviour and linguistic outcomes, but tend to force generalisations upon respondents. The present evidence on the relationship between vocabulary-related strategic behaviour and vocabulary gain indicates that strategies that involve repetition, including those that promote repetition through complex elaboration, are positively related to recall. Moreover, there is evidence to suggest that the more frequent use of a wider range of strategies leads to vocabulary gain.

2.4 INFORMAL L2 CONTACT, VOCABULARY-RELATED STRATEGIC BEHAVIOUR AND VOCABULARY GAIN IN A STUDY ABROAD CONTEXT

Exhaustive searches of major applied linguistics databases such as ERIC indicate that to date there have been no studies on the relationship between informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context. Therefore, the empirical evidence upon which the present study is based is hereby discussed in four sections: Firstly, the evidence pertaining to vocabulary acquisition in a study abroad context is detailed and evaluated. Then is discussed the present evidence on strategic behaviour in a study abroad context. Next is detailed and critiqued the one empirical study that has explored L2 contact, strategic behaviour (though not specifically vocabulary-related strategic behaviour) and vocabulary gain in a study abroad context; that of Segalowitz, Freed, Collentine, Lafford, Lazar and Díaz-Campos (2005). Finally, a summary of the evidence on informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context is drawn from the studies discussed in this section, into which the present study is situated.

2.4.1 VOCABULARY GAIN IN A STUDY ABROAD CONTEXT

As mentioned in 2.1.16, all of the studies conducted to date on lexical development as a result of SAE have found evidence of gains. Studies of lexical development have either taken a prospective, within-subjects design or a between-groups methodology that compares the gains of a study abroad group with participants in other settings. The evidence derived from both methodological approaches is now discussed.

2.4.1.1 Within-subjects studies of vocabulary gain in a study abroad context

In an investigation of the impact of SAE on rate of vocabulary acquisition, Milton and Meara (1995) employed a sample of mixed L1 participants (German, Italian, Spanish and French) who were studying on a six-month academic exchange at a British university. They administered the Eurocentres Vocabulary Size Test (Meara & Jones, 1990) pre and post the SAE. Each participant's pre-test results were divided by the number of years they had been studying English, thus providing the annual vocabulary acquisition rate of at home (AH) study. The researchers then compared this figure with the SAE vocabulary rate to find that the SAE rate was four times that of the AH rate. The participants who had the lowest vocabulary scores before the SAE were those who made the highest gains.

Another longitudinal study of the impact of SAE and proficiency level on lexical development was that of Ife, Vives Boix and Meara (2000). The researchers set out to measure 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. The sample was divided based on their proficiency at the outset of the SAE (intermediate and advanced), yet the researchers fail to report how proficiency was measured: quite an omission given that proficiency was their independent variable of primary interest. The researchers predicted, based on the findings of Milton and Meara (1995), that proficiency would have an impact on lexical development.

The pre and post SAE vocabulary measure administered by Ife *et al*. (2000) was a version of the *Three Word Association Test* (Vives Boix, 1995), which the authors claimed is a test of "improvements in lexical organisation, rather than simply...number of items acquired" (2000:3). The test presented the test-taker with 40 clusters of three L2 words each. In each cluster, two of the words were strongly associated with one another: the respondents' task

was to select the 'misfit'. Additionally, at both time points the sample were required to provide L1 translations of the 120 lexical items that comprised the clusters in the Three Word Association Test, and this was used as the measure of vocabulary size.

Both proficiency groups in Ife *et al*. (2000) made a statistically significant gain in both vocabulary measures between time 1 and time 2 ($p<.013$ in all cases) and, conversely to the findings of Milton and Meara (1995), there was no statistically significant difference between the gains of the two groups. Length of stay was positively correlated with gains in lexical organisation as measured by the Three 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 finding should be considered with some caution. No correlation was found between the "social and motivational experiences" of the sample during the SAE and their lexical gain (Ife *et al*., 2000:15); however, no conclusions can usefully be drawn from this finding because the researchers failed to report how these 'experiences' were measured.

In a recent qualitative study, Fitzpatrick (2012) explored in detail the vocabulary acquisition of one L1 Chinese learner of L2 English, who was on an academic SAE in the UK, in order to try to chart the incremental changes in his knowledge and use of individual lexical items. She used the Lex-30 test (Meara & Fitzpatrick, 2000), which operates similarly to a word association task, in order to elicit a L2 corpus from the participant, which was subsequently analysed in terms of the frequency levels of the vocabulary elicited. Over a period of eight months, Fitzpatrick (2012) found that, although her participant's use of collocations, association and derivational affixes improved in a linear fashion, his acquisition of individual lexical items did not: the participant's receptive and productive knowledge of spoken and written form and of form-meaning connections were subject to frequent, non-linear change. Although the findings are ungeneralisable, this

study provided insightful fine-grained detail about the development of depth of vocabulary knowledge as a result of a stay abroad.

More recently, Pizziconi (2013) set out to investigate the impact of the timing of SAE on the lexical development of n=48 L2 learners of Japanese at a UK university, who represented a variety of L1s and nationalities and whose degree course included a stay in the target language community. One group of participants went to Japan for the second year of their degree (SA2) and the other group went for their third year (SA3). Before the year in Japan the SA2 group received intensive Japanese instruction of a 'beginner' syllabus and the SA3 group a more extensive approach which covered the beginner syllabus and part of an 'intermediate' syllabus. It is not stated whether these syllabi labels refer to the general L2 proficiency level of the participants. There was no control for formal L2 instruction on the SAE, nor for informal L2 contact. Pizziconi (2013) administered a vocabulary test which she developed based on the vocabulary section of a Japanese language proficiency test. The measure tested 46 L2 items across four multiple-choice formats: either (1) gap fills; (2) morpho-syntactic judgement tasks; (3) matching an item to a definition; and (4) matching an item to synonyms. None of the tasks involved use of the L1. The author states that "with the exception of gap-filling, these tasks can be characterised as testing passive vocabulary knowledge" (2013:5): however, the gap-filling task required the respondents to write in the gap one letter (a-d) to represent which of the four words/phrases that they thought best completed the sentence. Therefore, all of the tasks were tests of receptive knowledge.

The vocabulary test in Pizziconi's (2013) study was administered at three time points: immediately before the SAE for each group, immediately after for each group, and 1 year after the end of the SAE for each group. Due to participant mortality, only n=27 (SA2=17, SA3=10) participants completed both the time 1 and time 2 vocabulary test and only eight

in each group completed the delayed post-test. The pre-test revealed that SA3 group knew more Japanese words than the SA2 group. A 2 x 2 ANOVA that took group (SA2; SA3) and time of testing (pre-test; immediate post-test) as the independent variables revealed that both groups made lexical gains from time 1 to time 2 [$F(1, 25)=118.13$, $p< .001$] and that the SA2 group made more gains than the SA3 group [$F(1, 25)=7.79$, $p=.01$]. Similarly to Milton & Meara (1995) therefore, those participants with a lower initial level gained more as a result of the SAE. No significant differences were found with regard to the timing of the SAE.

2.4.1.2 Between-groups studies of vocabulary gain in a study abroad context

Comparing the lexical gains of EFL learners of English in Israel (n=79) and ESL learners in Canada (n=103) respectively, Laufer and Paribakht (1998) focused on vocabulary knowledge in three dimensions: (1) receptive knowledge, measured using Nation's (2001) Vocabulary Levels Test; (2) controlled productive knowledge, measured using the productive version of the levels test (Laufer & Nation, 1999); and (3) free productive knowledge, for which the Lexical Frequency Profile (Laufer & Nation, 1995) was employed. The findings suggest that there was a significantly larger gap between the passive and active vocabularies of the SAE learners than between those of their at home counterparts. Moreover, the researchers subdivided the SAE group based on how long they had been in the target language community (group 1=six months or less; group 2=seven months to two years; group 3=two year or more) to find that the gap between passive and controlled active vocabularies was smaller for learners who had been in the ESL context for more than two years than for learners who had been there for shorter periods of time. Unfortunately, no corresponding investigation of the EFL group (i.e. to determine whether number of years of instruction would yield similar findings) was carried out, so whether this finding is due to length of SAE is ambiguous. Indeed, the long-stay SAE learners may

have had different types of L2 contact (e.g. relationships with native speakers) that might explain the smaller gap between their receptive and productive lexical knowledge.

In a study of the impact of a semester-long SAE on lexical development, Collentine (2004) divided his sample into two groups; one comprised of at home L1 English L2 learners of Spanish at a US university (n=20) and the other of SAE learners of Spanish in Alicante, Spain (n=26). His sampling frame controlled for out-of-class contact with Spanish and amount of formal Spanish instruction. Baseline proficiency level testing indicated that both groups operated at an intermediate level. The researcher measured the pre and post treatment vocabulary knowledge of both groups using the ACTFL Oral Proficiency Interview (OPI) protocol. Two two-minute samples from the pre and post interviews were taken and these 4 minutes of data from each time-point were analysed with reference to vocabulary use in seven word categories: conjunctions, pronouns, prepositions, adverbs, adjectives, verbs and nouns. A calculation was made of how many words from each of the seven categories each participant used per 1000 running words, and these calculations formed the basis of a comparison of the two groups' lexical development from time 1 to time 2. Collentine stated that the latter four lexical categories represent a learner's "core lexical base" (2004:234), whereas the first three are indicators of discourse coherence. The SAE group performed no differently from the AH group except for producing in the post-test a higher number of adjectives than did the AH group. Collentine (2004) himself states that a possible confounding variable in his study was the relative ability of his participants in terms of oral fluency; that is to say, more fluent participants have greater opportunity to exhibit lexical range.

In a similar approach to Milton and Meara (1995), Dewey (2008) compared the vocabulary growth of three groups of English learners of L2 Japanese: at home (n=22), domestic immersion (n=14) and SAE (n=20). Similarly to Laufer and Paribakht (1998) he employed

three vocabulary measures to test this growth: a picture matching test, a situational vocabulary test and a version of the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993) that had been adapted for use with learners of Japanese. To measure amount and type of L2 contact Dewey (2008) administered the Language Contact Profile (Freed *et al*., 2004) and also employed weekly learner journals. The data on informal contact gleaned from these self-report instruments was intended to help ascertain whether a correlation existed between informal L2 contact and lexical growth. In terms of informal L2 contact Dewey (2008) found that participants who reported the most frequent social contact with Japanese benefitted the most from the SAE, yet the reliability of this finding, given that the LCP required participants to state how frequently they had engaged in certain situations up to as much as 11 weeks prior, is questionable.

In terms of vocabulary gain, Dewey (2008) found that the SAE group outperformed the at home group on all of the three measures used. They did not, however, outperform the domestic immersion group, except on the Situational Vocabulary Test: on this measure, the SAE group demonstrated significantly higher gains ($p<.05$), which suggests that their ability to "define words typically encountered in everyday situations" (2008:134) developed further than their counterparts in both groups. Unfortunately, the at home group in the study received less formal English instruction during the 'treatment' period than did the immersion and SAE groups, which may account for the lower gains reported by the at home group across all three measures. To validate Dewey's assertion that SAE is more effective for L2 Japanese vocabulary development than at home study, this study would need to be replicated with groups that received an equal amount of instruction during the treatment period.

Similarly to Collentine (2004), Sunderman and Kroll (2009) compared the vocabulary development of at home and SAE language learners, taking their sample from learners of Spanish at a US university. The researchers administered a picture naming task, a reading span task and a translation recognition task. They found, in contrast to the findings of Collentine (2004), that the SAE group were able to produce and comprehend lexis more accurately and at a faster rate than the AH group. The study took as the basis for one of its research questions the *External Cue Hypothesis*, which posits that there is a link between what a learner is able to comprehend and produce in the L2 and the external linguistic content (i.e. the L2 input they receive in their context of learning). As such, one could argue that the sample in Sunderman and Kroll's (2009) study should have been split into those whose context afforded them richer external cues (i.e. the SAE group) and those whose context afforded less rich L2 cues (the AH group). However, the authors (2009:84) state that "Of the 48 participants, 34 had not studied abroad", but also that, "Most of the participants had spent the previous academic semester and/or summer abroad". Therefore, whether any of the sample were non-SAE learners is unclear, and as such the findings of this study are somewhat compromised.

In a similar study design to Collentine (2004), Foster (2009) sampled Iranian undergraduate students of L2 English, some of whom were studying at home and some on a SAE, and compared the lexical production of these groups with each other and with a group of native speakers that functioned as a baseline. Lexical production was measured in terms of the D measure (Malvern & Richards, 2002), which drives from mathematical modelling of how new lexis is introduced into increasingly larger productive samples. Foster (2009) found that the SAE group was more similar to the baseline group than the AH group in terms of the D measure and of lexical usage. In other words, the types of

words and phrases used by the SAE group were more native-like than their peers who had studied at home.

In their study of n=21 L1 Spanish learners of English in the UK, Llanes and Muñoz (2009) found that over a period of a 3-4 week stay their sample made significant gains in the proportion of lexical errors in the oral production [t=4.513, df=23, p=.000]. However, the lack of a control group in their study means that these gains may have occurred without the stay abroad. Serrano, Llanes and Tragant (2011) investigated the differences in terms of lexical complexity in oral and written production between L1 Spanish learners of L2 English in three settings: intensive at home study (n=69) at a Spanish university; semi-intensive at home study (n=37) at a Spanish university; and SAE (n=25) at a UK university. The intensive at home group received 110 hours of English instruction over a period of four and a half weeks; the semi-intensive group received the same number of hours of instruction over a period of eleven weeks; the majority of the SAE group had 8-12 hours per week of classes in English (including English classes). Lexical complexity was measured through analysis of pre and post-tests of written composition and oral narrative. The pre-tests were completed before the start of the respective courses. The post test data was collected with the intensive at home group fifteen days after the pre test and with the semi-intensive group two months after the pretest, which meant that both groups completed the post test after eighty hours of instruction. The SAE group completed the post-tests twice: once after fifteen days (for comparison with the intensive at home group) and again after two months (for comparison with the semi-intensive at home group). No significant differences were found between the intensive at home group and the SAE group, but there was a statistically significant difference on lexical fluency development in favour of the SAE above the semi-intensive at home group in both written [F(1,

60)=19.62, $p<.001$, partial $\eta^2 = .260$] and oral production [F(1, 35)=4.32, $p=.046$, partial $\eta^2 = .122$].

One possible problem with the study design of Serrano, Llanes and Tragant (2011) is that there was no control for informal L2 contact. This omission is odd, particularly in light of the researchers' comment that "L2 learning contexts vary in terms of the...opportunities they offer for...output and interaction with native speakers" (201:133). It may be the case that the differences found between the SAE and semi-intensive at home groups were due to differences in the quality and quantity of informal L2 contact they experienced over the space of two months. It may also be the case that the lack of difference between the SAE group and the intensive at home group in terms of lexical complexity can be explained by their having similar degrees of L2 contact. That is to say, it is highly possible that SAE learners do not experience large amounts of meaningful informal L2 contact within their first fifteen days in the target language community, and as such, the SAE group's contact with the L2 may have been very similar to the intensive at home group's between the pre and post tests.

One possible problem with comparing academic settings in a between-groups design is that the setting may not be the only difference between the groups: one could argue that the formal L2 input that AH and SAE learners receive is very different, particularly if the major of the SAE learner is not the L2 language, i.e. on an ERASMUS SAE, the learner may not necessarily be taking in-sessional or supplementary English classes, whereas the AH academic English leaner is likely to be, and thus a comparison between them in terms of vocabulary acquisition may be confounded as a result.

2.4.1.3 Summary of vocabulary gain in a study abroad context

There is firm evidence of a positive impact of SAE on vocabulary gain: repeated measures designs indicate that SAE yields significant gains in terms of size and organisation, and accelerates L2 vocabulary acquisition. Most of the between-groups evidence is in favour of SAE over the at home context with reference to vocabulary acquisition, although the evidence in support of SAE over an intensive at home language learning context is less persuasive. In summary therefore, there is a strong body of evidence that indicates that SAE is beneficial to the vocabulary gain of L2 learners.

2.4.2 STRATEGIC BEHAVIOUR IN A STUDY ABROAD CONTEXT

As mentioned in section 2.1.1.8, there is little evidence on whether and how SAE impacts upon language learning strategies, but what qualitative data does exist suggests that SAE has no significant impact on the strategies a learner employs (Huebner, 1995; Miller & Ginsberg, 1995). There is some qualitative evidence to suggest that learners are wont to recreate in informal L2 contact the interactions they engage in in formal language learning settings, as illustrated by the case study conducted by Carson and Longhini (2002) of a learner of Spanish on a SAE in Argentina. Qualitative research has moreover suggested that SAE learners may expect native speaker interlocutors in the study abroad setting to act as a language teacher behaves in the classroom, and may become disheartened to find that those they interact with are not necessarily interested in helping them learn (Miller & Ginsberg, 1995; Wilkinson, 1998).

In another qualitative approach to strategic behaviour and SAE, Briggs (2014) sampled n=95 mixed L1 learners of English who were on a SAE in the UK. She used simulations of commonly encountered informal L2 contact scenarios as stimuli in semi-structured interviews to elicit the vocabulary-related strategies that the sample used in response to

newly encountered L2 words. Briggs (2014) analysed the interview transcripts with reference to a self-developed taxonomy of the vocabulary-related strategies that are available to learners outside of the classroom. She found that each simulated scenario engendered very different strategic reactions from the sample and concluded that the physical and social restrictions of each individual informal scenario play an important role in the strategic behaviour that language learners manifest. As Briggs (2014) did not conduct the same interviews with at home learners, one cannot determine whether the strategic behaviour manifested by her sample in the simulated scenarios was particular to learners who have experienced study abroad, or is shared by learners across different contexts.

In terms of quantitative explorations of SAE and strategic behaviour, there is as yet scant and unconvincing evidence of a relationship between the two. Kojic-Sabo and Lightbown (1999), in their cross-sectional study of n=47 ESL learners of English in Canada and n=43 EFL learners of English in Yugoslavia, set out to determine whether there were differences in the vocabulary learning strategies of their sample based on their context of learning. 'Strategies' was operationalised as being how self-initiated their approach to L2 vocabulary was, and to what extent they used out-of-class time to focus on vocabulary acquisition. They administered a questionnaire based on the work of Sanoui (1992) to determine the sources of the participants' newly encountered vocabulary, the opportunities they had to practice these words, and their preferences with regard to working with newly encountered words. The questionnaire required participants to indicate their position on a cline (from self to teacher initiated) for each questionnaire item. They found that there was a significant difference between the groups in terms of their learner independence: the ESL group reported using more self-initiated strategies than the EFL group [$F(1, 88)=5.97$, $p<.05$]. Furthermore, there was a significant difference between the groups on their use of

strategies to review newly encountered words: the EFL group used more review strategies [$F(1, 88)=5.32$, $p<.01$] than the ESL participants.

What do the results of Kojic-Sabo and Lightbown (1999) reveal? Their operationalisation of the term 'strategy' in this study appears to have been conflated conceptually with informal L2 contact. For example, they include as strategy items in their questionnaire activities such as "listening to the radio" and "watching English news" (1999:179). As such, the only confident conclusion that can be drawn from their findings with regard to strategic behaviour might be that there was a difference between the groups on how self-initiated their contact with English that had the potential to lead to vocabulary acquisition was, and that the ESL learners were more self-initiated in this manner. However, one could argue that the ESL group, due to living in the target language community, had more opportunity provided to them by their very context in this regard (e.g. they may get into a taxi and listen to the radio station the taxi driver is playing, but this is not neither teacher nor self initiated behaviour). As such, the findings of Kojic-Sabo and Lightbown's (1999) study shed little light on strategic behaviour in a study abroad context.

Lafford (2004) conducted a study of the impact of SAE on the communication strategies used in oral interaction by n=46 L2 learners of Spanish, n=26 of whom embarked on a SAE in Spain. Lafford (2004:208) analysed OPI and role-play transcripts and coded the sample's strategic utterances into 26 separate strategies in three categories: direct (i.e. non-interactional) versus interactional, problem-orientedness (i.e. did the communication problem stem from a lack of resources, from the participant, or from the interlocutor?), and L1-based (e.g. codeswitching) versus L2-based (e.g. paraphrasing). Both groups significantly decreased the number of strategies they used from time 1 to time 2, but there was no significant difference between the groups' reduction in strategic behaviour. However, at time 2 there was a significant difference between the groups in terms of the

number of strategies used, with the SAE group using fewer than the AH learners [F(1, 43)=28.031, *p*=.000] in all three strategy categories. However, it is not clear whether Lafford (2004) also conducted an ANCOVA on the time 1 data from the two groups, and such it is not clear whether this difference in the sample existed before the SAE occurred.

In another quantitative study of strategy use in SAE, Adams (2006) studied the impact of a two to four month SAE on the strategic behaviour of American learners of a variety of L2s in a range of host countries, such as Brazil, Austria and France. She administered the Strategies Inventory for Language Learning (SILL; Oxford, 1990), a questionnaire designed to measure how far learners identify with six types of language learning strategies (metacognitive, cognitive, memory, affective, compensation and social), pre and post the SAE. All participants reported using learning strategies more frequently at time 2, and a positive correlation was found between self-reported gains in proficiency and self-reported gains in identification with strategies. However, as the SILL forces respondents to generalise their strategic action across situations (Winne & Perry, 2000; Tseng, Dörnyei & Schmitt, 2006) and proficiency was measured via self-report, whether these findings are an accurate portrayal of the impact of SAE on strategic behaviour is as yet not established.

2.4.2.1 Summary of strategic behaviour in a study abroad context

The data to date on strategic behaviour in a study abroad context suggests that there is an impact of SAE on strategic behaviour, but the data has not yet succeeded in defining the nature of that impact. Qualitative investigations have revealed an interplay of factors in the SAE setting, such as formal language learning experiences, physical constraints and contextual intention, which can influence the strategies a learner manifests. Quantitative studies have suggested that a SAE serves to increase the frequency with which strategies are deployed by a language learner. Studies have probed the use of vocabulary-related

strategies to find that SAE learners are more self-initiated, and communication strategies to find that SAE learners use fewer than their at home counterparts, yet problems with operationalising the key term and with reporting of statistical analyses bring into question the reliability of these findings. In summary therefore, there is little firm evidence to date on the specific strategies that learners are using in study abroad experience.

2.4.3 INFORMAL L2 CONTACT, STRATEGIC BEHAVIOUR AND VOCABULARY GAIN IN A STUDY ABROAD CONTEXT

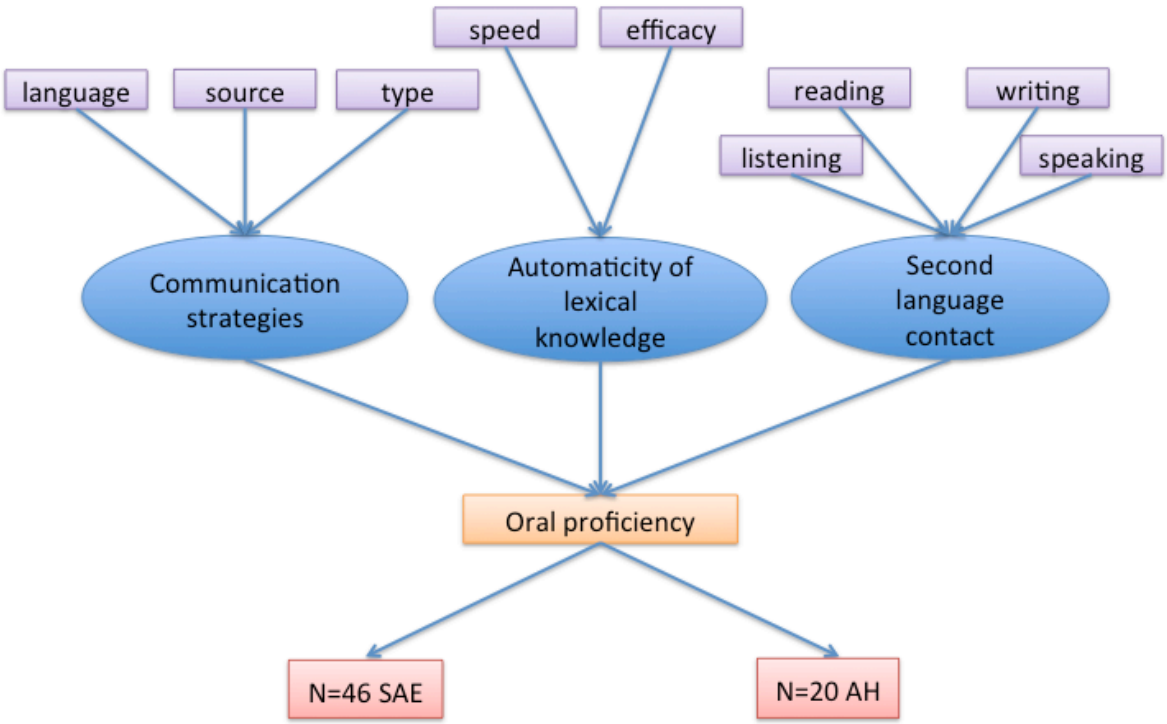
Segalowitz, Freed, Collentine, Lafford, Lazar and Díaz-Campos (2005) conducted the only empirical study to date that has explored informal L2 contact, strategic behaviour and vocabulary gain in a study abroad context. However, it focused on communication strategies as opposed to vocabulary-related strategic behaviour. The study took a between-groups design that compared the linguistic gains of n=46 North American L2 learners of Spanish who were studying either in Colorado, U.S.A. (n=26) or Alicante, Spain (n=20). A number of pre and post-test measures were administered at the beginning and end of one academic term of thirteen weeks in length. The measures relevant to the present study were: (1) the LCP (Freed *et al*., 2004); (2) the ACTFL OPI; and (3) a word recognition task, in which the participants saw L1 and L2 nouns on a computer screen and indicated whether each was animate or inanimate.

The data yielded from the word recognition test was analysed in terms of efficacy and speed of lexical access, and thus related to automaticity of lexical knowledge. The OPI data was analysed in terms of speech rate, mean length of run free of pauses, proportion of filled pauses and proportion of silent pauses of 400 milliseconds or longer. Strategic behaviour was operationalised as the communication strategies used by the sample to recover from breakdowns or gaps in the communication between participant and

interviewer during the OPI data collection procedure. The LCP data was analysed to yield a single score for each skill (reading, writing, listening, speaking) to represent the relative amount of time spent in each type of L2 contact. The design of Segalowitz *et al*.'s (2005) study is illustrated in figure 2.3.

The findings with reference to strategic behaviour revealed that 26 different communication strategies were reported by the sample, the most frequently reported of which were self-repair, restructuring, and accuracy checks. The researchers reduced their strategic behaviour data by categorising strategic instances into three dimensions: (1) language, which pertained to whether the strategy involved predominantly the L1 or the L2; (2) type, meaning whether the strategy was direct (e.g. the participant reformulating their utterance) or interactional (i.e. the participant involved the interviewer); and (3) source, which meant the strategy derived from either a resource deficit (a lack of knowledge on the part of the participant), an 'other performance' problem (meaning that the participant perceived a problem in the interviewer's speech) or a learner-focused problem (i.e. the participant has perceived a self-derived problem). The SAE group used significantly fewer strategies than the at home group at the post-test across all dimensions and the researchers report that ANCOVA revealed that this difference held true when pre-test strategic behaviour had been controlled for. Unfortunately, Segalowitz *et al*. (2005) did not report the significance value of this finding.

Figure 2.3: Study design of Segalowitz *et al*. (2005)



The findings with reference to vocabulary gain revealed speed and efficacy of lexical access were both positively correlated (p<.05) with oral proficiency in terms of mean length of run free of pauses. Both groups made significant gains from time 1 to time 2 on speed and efficacy of lexical access, but there was no significant difference between the gain sizes of the two groups. Participants who scored more highly in the lexical access test were those who gained the most in terms of oral fluency. No significant effects were found with reference to context of learning. The findings with reference to informal L2 contact revealed no significant correlations with oral proficiency gains. However, for the study abroad group the overall time spent in informal L2 contact was significantly negatively correlated [r=-.469, p=.016] with the use of communication strategies at the post-test (Lafford, 2004), suggesting that the more informal contact with the L2 a participant had during the SAE, the fewer communication strategies they used. This finding is particularly interesting in light of the lack of significant correlation between strategic behaviour and

oral proficiency because it suggests that the fewer communication strategies manifested by the SAE group at time 2 may not be as a result of fewer breakdowns of communication.

The findings of the study conducted by Segalowitz *et al*. (2005) here discussed indicate that study abroad experience decreases the number of communication strategies a learner uses, but this finding may be an indirect result of the significant gains in oral proficiency made by the SAE group. The findings also suggest that SAE has no more impact on automaticity of vocabulary knowledge than an at home context. Unfortunately, as oral proficiency was the dependent variable in Segalowitz *et al*.'s (2005) study, the researchers did not analyse whether there was any relationship between strategic behaviour and vocabulary gain, or informal L2 contact and strategic behaviour. Therefore, other than to indicate a possible relationship between L2 contact and communication strategies, the findings of Segalowitz *et al*. (2005) shed little light on the relationships between informal L2 contact, strategic behaviour and vocabulary gain in a study abroad context.

2.4.4 SUMMARY OF INFORMAL L2 CONTACT, VOCABULARY-RELATED STRATEGIC BEHAVIOUR AND VOCABULARY GAIN IN A STUDY ABROAD CONTEXT

As we have seen, there is a strong body of evidence that indicates that SAE is beneficial to the vocabulary gain of L2 learners. There is little evidence to date on the specific strategies that learners are using in study abroad experience or their relationship to vocabulary gain in the SAE context, yet the body of evidence on strategy use and vocabulary gain more generally indicates a strong positive relationship between the two. Very little evidence exists on the relationship between informal L2 contact and linguistic gain on a SAE and what does exist appears to suggest that there may be negative relationship between the two. However, as the vast majority of studies into SAE that have included informal L2

contact as a variable have measured the construct via frequency-based retrospective self-report, we cannot confidently rely upon their findings. To date, no studies have examined the relationships between informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context. In conclusion therefore, three main gaps in knowledge arise from this analysis of the literature. Firstly, the research into informal L2 contact has tended toward either (1) descriptive case studies that focused on very small numbers of participants, or (2) larger scale studies that required participants retrospectively to state the frequency with which they had encountered certain L2 contact scenarios. As such, there is a dearth of reliable L2 contact data from which generalisations can be made. Secondly, research into vocabulary-related strategic behaviour that occurs outside the boundaries of the laboratory or classroom has for the most part required respondents to generalise their behaviour across contexts and/or tasks. As a result, the evidence to date on vocabulary-related strategy use in informal L2 settings must be considered cautiously in terms of its reliability and validity. Finally, no present evidence exists on the interplay between informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context.

CHAPTER III

METHODOLOGY

This chapter comprises the key research questions, an overview of the methodological approach, the study design, and the data analysis approach.

3.1 RESEARCH QUESTIONS

This study sought to address the following research questions:

1. Is there a relationship between identification with informal L2 contact scenarios and vocabulary gain in a study abroad context?

    1a. Do location of SAE and length of stay play a role in this relationship?

2. What vocabulary-related strategies do learners of English employ in informal L2 contact in a study abroad context?
3. Is there a relationship between vocabulary-related strategy use in informal L2 contact and vocabulary gain in a study abroad context?

    3a. Do location of SAE and length of stay play a role in this relationship?

The three main research questions are presented in this order to match the data collection and analysis procedures of the study.

3.2 STUDY DESIGN

The four-phase, longitudinal, embedded, mixed-methods design of this study followed the typology:

QUAN → QUAN → QUAN + qual → QUAN

That is vocabulary testing at two time-points, and within those two time points a questionnaire and the use of computer-based simulations as a stimulus in semi-structured interviews. Embedded mixed-methods designs derive from the tenet that a single method is insufficient to answer one or more research questions (Creswell & Plano Clark, 2007) and this kind of design embeds a supportive method into the other, principal method (Punch, 2009). In this study the principal method was that of the QUAN research tradition, with strategies from the QUAL tradition embedded to supplement and enhance the numeric data with reference to context-specific strategic behaviour. The design was driven by a deductive theoretical thrust and was selected based on a pragmatic, 'fitness for purpose' basis with reference to the three principal research questions.

The first research question relates to the relationship between identification with informal contact scenarios as the independent variable and L2 vocabulary gain in a study abroad context as the dependent variable. Research question one aimed to test the hypothesis that there is a positive correlation between informal contact and vocabulary gain in a study abroad context. It is widely held that informal contact in a study abroad context has a positive impact on proficiency gain in general (Kinginger, 2009; Tanaka & Ellis, 2003), but scant evidence exists to support the specific influence of informal contact on L2 vocabulary gain.

A quantitative approach to research question one was appropriate because it required the use of a sample large enough to be representative of the population, and numeric data to demonstrate with statistical significance the effect of the independent variable on the dependent variable. To answer this question, QUAN data from the time 1 and 2 vocabulary tests was analysed in light of the QUAN data from part 1 of the questionnaire (which measured identification with different types of informal L2 contact). A sample size of n=241 participants was arrived at for both the time 1 vocabulary test and the questionnaire.

In a previous study (Briggs, 2010), participant mortality between testing at time 1 and 2 stood at approximately 20% and in the present study it stood higher than this at 23%, thus leaving a sample size of n=196 at time 2.

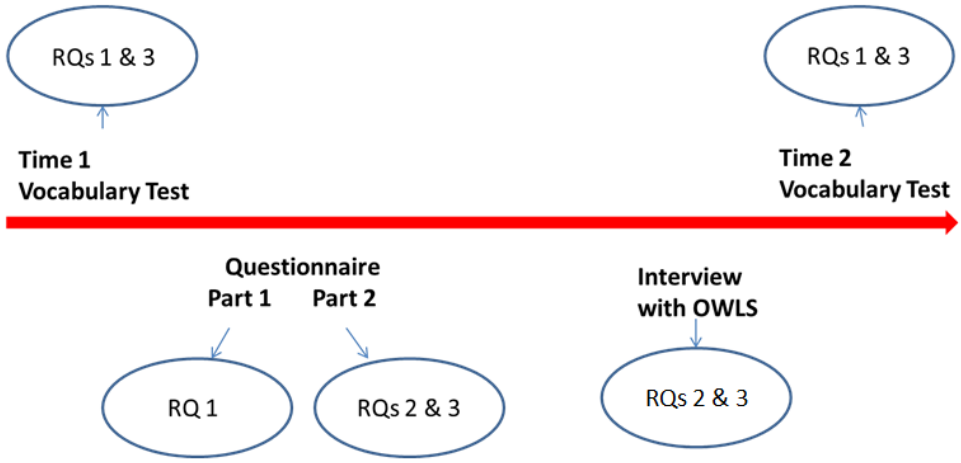
The second research question addresses the types of vocabulary acquisition strategies employed in informal contact with the L2 in a study abroad setting. This question was addressed quantitatively and qualitatively at both the data collection and analysis stages. Initial QUAN data on strategy use was taken from section 2 of the questionnaire (which measured identification with vocabulary-related strategies), and analysis of this data enabled generalisations to be made about which strategies are identified with more or less by learners in a study abroad context. Further QUAN data was collected and analysed from the interview transcripts. This dual data analysis approach was employed to allow a multi-level analysis which answers not only the ‘what’, but also the 'when'; a window into how different contexts affect the use of strategies/combinations of strategies. Supplementary QUAL data to enhance the revelatory power of the simulation tool was taken from the interview transcripts, and used to provide depth and richness to the data on strategy use.

Research question three probes the relationship between out-of-class strategy use and vocabulary gain in a study abroad context. Research question three aimed to test the hypothesis that the use of a greater quantity and variety of vocabulary-related strategies (including strategic combinations) has a positive impact on vocabulary gain (Lawson & Hogben, 1996; Macaro, 2003) in the study abroad context. To answer this question, QUAN data from section 2 of the questionnaire (identification with strategic behaviour), from the vocabulary tests at time 1 and 2, and from the transcripts of the interviews were analysed. Additional insight into this question was gained using QUAL data from the interviews (comments the participants made about strategy use and its perceived efficacy), shedding

light on similarities in strategic behaviour between those participants who were most successful at gaining vocabulary and those who were not.

Figure 3.1 shows the data collection stages in sequence and the research question(s) to which the data from each stage pertains.

Figure 3.1: Data collection sequence



3.3 POPULATION

The population is study abroad learners in England, learning English as a second language. They are adults (over eighteen years of age), studying 'general English' (i.e. preparation for everyday situations and language in English) at private language institutions (rather than taking part in an academic exchange such as ERASMUS). The population undertakes a minimum of fifteen hours per week contact time of English language instruction (as per the UK regulation for approval of a General Student Visa). They spend the majority of their time on the SAE outside of the classroom: Although there are no sources to support this claim directly, the majority of English language institutions (e.g. Bell English, 2014; Embassy English, 2014) state on their websites that study abroad General English courses provide between 16 and 23 hours per week of formal English language instruction over the five day working week. If a fourteen-hour waking day is presupposed, then the SAE

learner spends 98 waking hours per week in the SAE and as such the average number of waking hours spent in language classes during a SAE comprises less than one quarter of their overall waking time. Five years' experience of teaching and researching study abroad learners has informed the present researcher as to how this population spends their time outside of class on a SAE: working a part-time job, volunteering, at-home study, socialising, and travelling around the UK and Europe.

The length of stay in the host country is a minimum of 6 weeks. The population comprises a wide variety of nationalities and L1s, and they study in multilingual classes in England. They are predominantly in their 20s and 30s, and their reason for coming to England to study English is to increase their job prospects in their home country, for the enjoyment of learning English as an L2, and/or for experience of living abroad. They are living either with a host family, in dorm-style accommodation arranged by their language school, or in private rented accommodation.

3.4 SAMPLE

The sample (n=241) was taken from students of language institutions in Oxford and London because these are highly popular destinations in England for study abroad learners to enroll in language courses. Of the 241 participants, n=112 (46.5%) came from four London-based institutions and n=129 (53.5%) from three Oxford-based institutions. By choosing two cities of very different sizes and cultural makeups, the researcher hoped to ascertain whether location affects the type and/or amount of informal contact with the L2, and indeed the strategic behaviour applied in that contact. E.g. in a larger, more multicultural city such as London, a larger number of SAE learners may reside within their L1 community or interact more with L2 speakers of their L1, and thus have a different

experience of informal contact with English (and develop differing strategies to manage that experience) than those elsewhere.

To investigate the extent to which the sample participants were representative of the population of the study (i.e. study abroad learners in England), the researcher contacted ten private language institutions (which did not participate in the present study) from four cities in England (Cambridge, Manchester, Stratford-upon-Avon and Bournemouth) to ask for approximate demographic information about their student bodies. According to the data obtained from these institutions, a match between sample and population was achieved through an equivalent match of nationalities, L1s, hours of formal instruction per week in the UK, and UK living arrangements. Other than numbers of students enrolled, there were no discernable differences between the student bodies of the Oxford and London schools from which the sample were taken and other private language institution student bodies across England.

In this study there was no restriction on nationality or L1 to allow for the possibility that data analysis might show key differences in amount and type of informal contact and strategic behaviour based on these factors. Participants had no prior SAE because experience of study abroad might have impacted their engagement in and strategic responses to L2 contact outside of the classroom. The sample comprised participants receiving a minimum of 15 hours per week and maximum of 25 hours per week of general English instruction. Their proficiency level was a minimum of the Common European Framework of Reference level B1 (low-intermediate) to ensure they had the linguistic ability to complete the questionnaire and potentially take part in the interview.

To participate, a learner was required to be undertaking at least a six-week SAE in order to have time to participate in all three data collection stages. The previous studies (Briggs,

2010, 2014) suggested that the average SAE (in Oxford) is less than 20 weeks; therefore the longest period of participation in the present study was 20 weeks. N=55 participants were in the short stay group (6-10 weeks; 22.8%), n=56 in the medium stay group (11-15 weeks; 23.2%) and n=130 in the long stay group (16-20 weeks; 53.9%).

All n=241 participants completed the time 1 vocabulary test and the questionnaire. Due to participant mortality, n=196 completed the time 2 test. N=36 participants were interviewed using the Opportunities With Language Simulator (OWLS). The initial aim was to interview 40 participants and thus to have 20 matched pairs; however, due to participant mortality, data from only 18 matched pairs was gathered. To reduce the total number of participants from 241 to 40, the matched pairs were selected as follows: participants were grouped according to the percentile (e.g. 1st quartile, median, 3rd quartile, 4th quartile) into which their overall vocabulary gain scores fell. Similarly for identification with informal L2 contact scenarios and vocabulary-related strategic behaviour (from the questionnaire parts 1 and 2), a mean identification score was calculated for each participant and they were grouped in the same way. A matched pair of participants scored in the same quartile for overall vocabulary gain, for identification with informal L2 contact scenarios, and for identification with vocabulary-related strategies. The only difference between participants in a pair was their L1; one language which uses a Latin-based writing system and one which does not (the rationale for taking writing system as the matched pair difference is provided in section 5.1.2).

3.5 RESEARCH INSTRUMENTS

Four research instruments were developed for use in this study. These were a vocabulary test, a questionnaire, a computer-based out-of-class contact simulation tool (the Opportunities With Language Simulator, or 'OWLS') and an interview schedule for use

with the OWLS. Additionally, a taxonomy of out-of-class vocabulary-related strategies was developed for use as an analysis instrument and is discussed in 3.5.5.

3.5.1 VOCABULARY TEST

The vocabulary test (Appendix D) designed for use in this study was adapted from Nation's (1990) Vocabulary Levels Test. This diagnostic test is designed to reveal how many high-frequency words a learner knows and covers both receptive and productive vocabulary knowledge (the productive version having been developed by Laufer and Nation, 1995). It utilises vocabulary items at five frequency levels: 2000, 3000, 5000, AWL (Academic Word List) and 10,000. The words included in each level were established using word frequency data from a number of sources, including Thorndike and Lorge (1944) and West (1953).

The levels test was developed on the premise that learners of English 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 words and three definitions, so that three of the words function as distractors. As such, the receptive test is a test of passive recognition of the orthographic form-meaning link and of knowledge of meaning in terms of concepts, referents and associations. The productive test comprises 18 words at each frequency level. Each question is a sentence in which the target word is incomplete; the respondent 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). In both the receptive and productive parts, the use of nouns, verbs and adjectives as the target word is at a ratio of 3: 2: 1 respectively.

There exist many versions of the levels test: the original test was written by Nation in 1983 and re-published by the author in 1990; in 1993, Schmitt wrote a revision of the original test (version A) and also wrote versions A-D. Laufer and Nation (1999) then developed a productive ability vocabulary levels test. In 2001, Schmitt, Schmitt and Clapham created versions 1 and 2, which are amalgamations of receptive versions A-D. Additionally, there are a number of bilingual versions of the test available.

3.5.1.1 Validity and reliability

The levels test is widely used in both testing and research contexts across the globe (Schmitt, Schmitt & Clapham, 2001) and therefore a number of studies have sought to determine its validity and reliability. Read (1988) sought to validate the original version of the levels test, concluding that the test was reliable (Cronbach's alpha >.91) and that test scores by frequency level tended to adhere to an implicational scale, whereby each level could only be mastered if the preceding levels already had been. Beglar and Hunt (1999) investigated the 2000 and AWL levels of versions A-D to find evidence of convergence between scores on these levels and those of the Test of English as a Foreign Language (TOEFL). Moreover, they found that these levels were measuring a single construct and that there was a strong relationship between items in the sections under investigation.

Schmitt, Schmitt and Clapham (2001) took a mixed-methods approach to the validation of two new versions of the levels test using scores from n=801 participants of a wide variety of nationalities, L1s and proficiency levels. Version 1 comprised an amalgamation of versions A and B and version 2 of C and D. The researchers found high levels of reliability in both tests as measured by Cronbach's alpha: no frequency level in either version resulted in a value smaller than .92. Moreover, the researchers found a very high scalability value (.978) between the four frequency levels, thus substantiating Read's

similar (1988) finding and providing initial evidence in support of the validity of these tests. In addition, the researchers investigated the equivalence of version 1 and version 2, finding no statistically significant differences between the means and variance in scores at any level.

In an investigation of the reliability of the 5000 level of versions A and B, Xing and Fulcher (2007) concluded that although version B contains a greater number of more difficult items, both tests at this level are highly reliable and highly correlated with one another. Stewart and White (2011) sought to estimate the effects of guessing on scores in response to the levels test by examining the relationship between the number of words in a test (including the distractors) known by a respondent and any score increase due to guessing. The scholars reported that the more test words a respondent knows, the higher the probability that they will correctly guess any subsequent unknown items. It is difficult however to draw a firm link between this finding and the levels test itself: Stewart and White (2011) administered 100,000 self-made, 99-item version tests, which mirrored the multiple-choice format of the original levels test and versions A-D. Unfortunately, the researchers provide no information regarding their choice of vocabulary for use in these tests, nor of the match of ratio of word types to the original test(s), nor of the frequency levels their tests were designed to measure.

In a criterion referenced approach to investigating the validity of version B of the 2000, 3000 and AWL levels tests, Platzer (2012) took as his criterion measures the Oxford University Press' Quick Placement Test, the X-Lex vocabulary size test (Meara & Milton, 2003) and the receptive skills and grammar knowledge of his participants as defined by the CEFR descriptors for their respective proficiency levels (how participants' proficiency was measured is not described). Regarding scoring validity, Platzer (2012) found consistency estimates of between .80 and .88. In terms of discriminant validity, differences between the

correlations were found to be statistically significant (p=.01), and hierarchical cluster analyses revealed that the levels test scores converged on those of the X-Lex in terms of the link between vocabulary size and CEFR level.

In terms of the productive levels test, Laufer and Nation (1999) reported a Cronbach alpha coefficient of .77 for the 2000 level of version A and between .51 and .80 for the other versions at the same level. Laufer & Nation (1999) also investigated the discriminant validity of their productive test, finding the test well able to differentiate between learners of differing proficiency levels. Abdullah, Puteh, Azizan, Hamdan and Saude (2013) conducted a validation study of a 500-word productive level test that they created in order to adapt the levels test for use with low proficiency learners of English (i.e. learners whose vocabulary range fell below the 2000 level), taking version A of the 2000 level productive levels test as their criterion measure. The researchers found high levels of reliability for both tests with Cronbach's alpha coefficients of .63 for the 500 level, .72 for the 2000 word level, and .80 for the 500 and 2000 levels combined. There was a significant difference (p<0.05) between the mean scores on the 500 and 2000 level tests, with higher scores in response to the 500 level.

3.5.1.2 Test selection

The levels test was chosen as the basis for the vocabulary testing in this study for a number of reasons. Firstly, it covers both receptive and productive vocabulary, so provides 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 part of the 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 parts in terms of sensitivity to knowledge. Thirdly, unlike other widely-used measures of testing vocabulary breadth such as the Eurocentres Vocabulary Size Test (Meara & Jones, 1990), the levels test does 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 more accurate snapshot of actual knowledge than perceived knowledge, and is not affected by the test-taker's honesty. The levels test is an L2-only test that is a necessary feature in a study that employs a sample of mixed L1s.

Although the levels test is sensitive only to the 10,000 word level, and some of the participants in this study were deemed to be at C1/C2 level by their language institution, any possible ceiling effect (particularly at time 2) was not considered to be a threat because in piloting the test in a previous study (that of Briggs, 2010), learners of English at a much higher proficiency level (IELTS band 7.5) than the participants had not completely mastered the vocabulary at this frequency level. Ideally, a simple L2 - L1 vocabulary test would be adopted in this study, whereby the levels test definitions occurred in participants' L1 or the L1 equivalent of the target words could be selected; this would mean the responses would not rely on comprehension of the possible answers. However, due to the wide variety of L1s anticipated in this study, a L2 – L2 test was necessary.

For the purposes of this study, items at the AWL were excluded because none of the participants were undertaking the study of English at university level. The receptive part of the levels test has a version A and B. Items from version A only were used in this study

because version A was developed prior to version B, and has therefore been used more extensively. The eighteen target items per level in the productive test were decreased to ten items per level. This was in consideration of the amount of time the test would take to complete; much longer than forty minutes would be asking too much of the participants as they were volunteering their free time to complete the test. Some of the eight items omitted at each level were chosen because they were specifically American English words (e.g. 'diaper', 'sophomore') and therefore would be unlikely to be acquired in informal contact with British English. The others were selected at random by writing one word each on a slip of paper and then picking eight of the eighteen words out of a bowl. The omission of eight lexical items per frequency level may have impacted the validity of the test in terms of measuring a learner's vocabulary size; however, as the test was not used in this study to estimate breadth of knowledge but rather to indicate vocabulary gain (and in the same format in Briggs, 2010, was able to differentiate between the sample based on gain), the modification from eighteen to ten words per level was not deemed a threat to the validity of the present use of the test.

The two parts of the test were presented separately, and collated separately, in order to highlight the difference in task type between the two and to allow an example question with instructions on the first page of each part. In both parts, instead of presenting the items in order of word level (i.e. starting with all of the 2000 level words and progressing to the 10,000 level words), 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 levels at the beginning of the test to prevent the participants from becoming disheartened early on. Although the test was piloted in a prior study, it was piloted again in order to ensure its suitability to the requirements of the present study.

3.5.2 QUESTIONNAIRE

The questionnaire administered in the present study comprises three parts: Part 1 was designed to measure identification with different types of informal L2 contact; Part 2 was designed to measure identification with out-of-class vocabulary strategies; and Part 3 was designed to collect background data about the participants.

Part 1 comprises 28 items, each pertaining to one informal contact 'scenario' (e.g. 'I read novels in English outside of class' / 'I use English to get directions or information from strangers'). The items derived from the LCP (c.f. 2.1.2.3a). The LCP was designed to capture *amount* of informal contact in terms of the L2 input and use that learners of Spanish on a SAE experienced. The LCP comprises two sections (a pre- and post-'test' section; so called because they were designed for use before/at the beginning of a SAE and after/at the end respectively) in order to facilitate comparison between the informal L2 contact language learners experience in the L1 context as compared to their informal L2 contact whilst resident in the target language community (i.e. during the SAE). The LCP was chosen as the basis of Part 1 of the questionnaire because, other than with reference to technology-assisted L2 contact, the LCP offers comprehensive coverage of the different types of informal contact which study abroad learners may encounter. However, it was adapted, added to and amended so much so that the end result bore little resemblance to the original instrument. A summary of the key changes made from the LCP to the questionnaire administered in the present study is given below:

- All items were amended to pertain to learners of *English* as their L2.
- The pre- and post-SAE sections were amalgamated because the present sample had already begun the SAE at the point of data collection.
- The Likert scale was changed from a frequency rating to a 'how true of me' rating.

- Items that included the verb 'try' were reworded.
- Part 2 was added to collect data on identification with vocabulary-related strategies.
- Part 3 was added to collect background data about the participants.

The LCP's frequency-based response scale (a 5-point scale from 'Never' to 'Daily') was changed because it is far less reliable to gauge accurately how often something happens than to state how representative it is of one's experience. Similarly, in terms of responses to the items in Part 2 (which measure identification with vocabulary-related strategies), the frequency of a strategic action may be wholly unrelated to its efficacy (Dörnyei, 2003): that is to say, a learner may engage frequently in a strategic behaviour which has little to no positive impact on his/her L2 comprehension/acquisition. Furthermore, frequency scales of items of a different nature cannot be treated cumulatively in terms of QUAN analyses (Dörnyei, 2003). The Likert scale appears on every page of the questionnaire administered in the present study so that respondents' cognitive capacity was freed from the burden of having to remember which number (1-5) represents which 'how true of me' statement.

Some LCP items contain the verb 'try', e.g.:

How much time did you spend trying to catch other people's conversations? (2004:356)

These items were reworded to read, e.g.:

I listen to other people's conversations in English outside of class.

The difference between these statements is that the first is strategic (something done with the intention of enhancing L2 acquisition) whereas in the second the learner's intention

may or may not be to acquire language (i.e. the scenario may be naturally-occurring), but the scenario nevertheless represents an opportunity for strategic behaviour to occur.

In the LCP, items referring to spoken contact with the L1 are presented together, followed by reading, listening and writing respectively. In Part 1 of the questionnaire administered to the present sample, the questions were randomised so that when analysed using factor analysis, no pre-imposed grouping affected which items loaded onto which factor.

All Part 1 items in the questionnaire are worded in the positive. One potential disadvantage of including only positive statements as questionnaire items is that respondents may fall into a pattern of response rather than consider carefully how far they identify with each scenario. However, participants' clear comprehension of the items was considered the most important factor in their wording, and the calculation of Cronbach's alpha from the pilot data (.86) indicated that pattern response had not occurred.

A free-write section was added beneath the eight Part 1 items identified in a previous study (Briggs, 2010) as the informal L2 contact scenarios most highly identified with by the population under investigation in the present study. The number eight was arrived at because there were eight scenarios in the 2010 study that had yielded an overall mean of 2 or lower, meaning that these eight scenarios were positively identified with by the 2010 sample (because 1=This is very true of me; 5=This is not at all true of me). The free-write sections comprise probing questions designed to elicit details of participants' typical experience of each scenario. This is an example of intra-method mixing in MMR (Johnson & Turner, 2003); that is, the use of one data collection strategy (the questionnaire) which utilises both QUAN and QUAL factors, e.g. closed (i.e. yes/no, Likert scale) items and open (i.e. free-write) items. These details were used to inform the design refinement of the simulations of the most identified with scenarios in the OWLS instrument (c.f. 3.5.3).

Part 2 of the questionnaire comprises 28 items, each pertaining to one vocabulary-related strategy. This study adopted a cognitive operationalization of the term 'vocabulary-related strategy' based on Macaro's (2006) framework, whereby strategies are (1) conscious and goal oriented; (2) located in working memory; (3) discrete (i.e. they do not involve the use of other strategies); and (4) situation- and task-specific yet transferable to other situations and tasks. In terms of situation-specific behaviour, the strategies included in the questionnaire were selected based on their availability to the learner in informal L2 contact (i.e. contact with English that occurs outside of the classroom setting). The strategies derived from the taxonomies of Schmitt (1997), Nation (2001), Gu and Johnson (1996) and Oxford (1990). Strategies that are clearly classroom-based, e.g.:

Ask teacher for a L1 translation (Schmitt, 1997:7)

or cannot not be applied to lexical comprehension/acquisition were not included. Similarly, strategies that did not meet the stringent operationalization of the term in this study (e.g. were too global) were excluded.

The items which gathered background information about the sample (Part 3) were positioned at the end of the instrument in order to ensure that any fatigue provoked by completing the questionnaire was less likely negatively to affect participants' responses to the most cognitively challenging items (those requiring consideration of their out-of-class contact with the L2 and their strategic behaviour). Part 3 collects data on factors such as age; L1; previous SAE; medium of primary and secondary education (i.e. EMI, mother-tongue or L-other instruction); English instruction during the SAE; L2 proficiency level; and living arrangements during the SAE.

The questionnaire was administered to n=64 pilot participants who matched the population of the present study (i.e. they were adult SA learners of English who were studying at

private language institutions, had a low intermediate proficiency level and above, and who represented a range of nationalities and L1s). The internal consistency of the pilot data was calculated using Cronbach's alpha, providing a high coefficient of .86.

3.5.2.1 Validity and reliability

The internal consistency values for the informal contact and strategic behaviour sections of the questionnaire are reported in sections 4.2.1 and 5.1.1 respectively.

Very little statistical information on reliability of the LCP is available from searching of major linguistics databases: Briggs (2014) reports an alpha statistic of .85 for her version of the tool (administered to a similar sample to that of the present study). Martinsen (2011) adapted the LCP for use in a study of the impact of short-term SAE on cultural sensitivity in American learners of Spanish and reports a strong Cronbach's alpha coefficient of .83, yet the researcher fails to include in his article any specific details regarding how the instrument was adapted and changed from Freed *et al*.'s (2004) original. Despite its wide use in SLA research, no evidence has been reported regarding the validity of the LCP or any modified version of the LCP. In terms of the questionnaire that was administered to the sample of the present study, to achieve face validity all participants with whom the questionnaire was piloted were invited to comment upon the extent to which the items in Parts 1 and 2 were representative of the constructs they intended to measure (i.e. informal L2 contact on a SAE and strategic behaviour in informal L2 contact respectively). Furthermore, the pilot participants were asked if they would change, amend or add to the content of each part of the questionnaire in order better to represent the construct under investigation. Where suggestions were made, these were taken into consideration by the researcher and were enacted where relevant. In another approach toward construct validity, two expert judges (both professors of Applied Linguistics at high-ranking international

universities) appraised the instrument and deemed it to have a high level of content validity. With reference to criterion-related validity, the findings of the informal L2 contact and strategic behaviour data (reported in sections 4.2 and 5.2 respectively) provided strong evidence of concurrent validity in that the instrument was able to distinguish between participants who identified highly with the two constructs and those who did not.

3.5.3 THE OPPORTUNITIES WITH LANGUAGE SIMULATOR (OWLS)

The OWLS was developed by Briggs (2014) to simulate the five most identified with informal L2 contact scenarios as identified by the questionnaire administered in the 2014 study (table 3.1) and to be used as a stimulus to elicit vocabulary-related strategic behaviour in semi-structured interviews. Briggs (2014) chose to simulate five scenarios in order to keep the duration of the OWLS interviews to 40 minutes (i.e. she predicted that each simulation would prompt approximately eight minutes of discourse). To create each simulation, data from the free-write sections of the questionnaire (administered in the 2014 study) that referred to each scenario were analysed. The most commonly occurring details were used to inform what visual and audio factors to include.

Table 3.1: Scenarios simulated in the OWLS in Briggs (2014)

Simulation	Questionnaire item number	Scenario	Target word
1	6	Short exchanges in English with cohabitants	fellow
2	22	Listen to English DVDs/movies	nuisance
3	20	Read timetables, announcements, posters, menus etc.	urge
4	3	Speak in English to service personnel	bargain
5	25	Listen to English songs	vain

Each simulation comprises three PowerPoint slides (Appendix F): Slide 1 shows the number of the simulation. Slide 2 functions as the 'scene-setter', comprising a photograph with an accompanying background MP3 designed to situate the participant to the context

being simulated. Slide 3, the stimulus, shows a photograph of the same setting and/or interlocutors as slide 2, or illustrates what the interlocutors in slide 1 are looking at. If the scenario involves an interlocutor, an MP3 of a question/sentence is heard in slide 3. If not, the question/sentence appears in written form. Within each question/sentence there appears one 'target' word that had been predicted by the researcher to be unfamiliar to the participants, and in response to which they would have the opportunity to take strategic action.

The OWLS target words were chosen from the General Service List (GSL; West, 1953), which comprises 2,000 headwords deemed to be most useful to learners of English. The researcher selected eighteen words unlikely to be known by the sample (based on lengthy experience of English language teaching) yet possible to acquire through contact with the target language community. Half of these items came from the first 1000 words of the GSL, half from the second. The eighteen words were piloted with n=8 learners in the multiple choice format of Nation's (1990) Vocabulary Levels Test, a diagnostic test designed to reveal how many high-frequency words a learner knows. Each question in the levels test presents the test taker with six words and three definitions so that three of the words function as distractors. Because the six words are semantically very different, even partial word knowledge is sufficient to answer correctly (Beglar, 2000). This sensitivity complements well the out-of-class acquisition investigated in this study; outside the classroom it is less likely that learners 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. Eighteen words were chosen in case most of the pilot participants knew more than half the words; even if that were so, there would be sufficient words remaining to facilitate simulations of the top five scenarios.

The pilot revealed that the population were likely to be familiar with most of the words taken from the first 1000 of the GSL. Therefore only one of the words used in the OWLS came from this group, and the others from the second 1000. The final five target words were chosen because they were unknown to the learners with whom the lexis was piloted, and based on their applicability to the scenarios to be simulated.

Three simulations were added to the OWLS for use in the present study. These simulations each took a phrasal verb as the target vocabulary item presented in the stimulus. The decision to encompass multi-word items was taken for a number of reasons. Firstly, an instrument pertaining solely to single-word lexical items is not a comprehensive representation of the English language: Corpus linguistics reveals that a significant proportion of written and spoken English consists of multi-word items (Altenberg, 1998; Erman & Warren, 2000; Schmitt & Carter, 2004), and Lewis' (1993) Lexical Approach provided a persuasive argument for the teaching and learning of these in second language acquisition. There is evidence to suggest that "ordinary language experience" (Coady, 1997:282) does not provide adequate means for the acquisition of multi-word items. If that experience encompasses informal exposure to the L2 in a study abroad context, then additional and context-rich multi-word input is a benefit to the learner. Moreover, there is evidence to suggest that despite their frequency and prominence, second language learners of English find phrasal verbs difficult to acquire (Moon, 1997; Darwin & Gray, 1999; Celce-Murcia & Larsen Freeman, 1999). This may be due to their multidimensionality with regards meaning (Gardner & Davies, 2007) and can result in circumvention of the problem by way of total avoidance of phrasal verbs (Liao & Fukuya, 2004; Hulstijn & Marchena, 1989).

A phrasal verb consists of a verb ('go') and one or more adverbial particles ('over'/'out with') which together comprise one lexical and syntactic unit; lexically singular because

neither item alone conveys the meaning of the two together, and syntactically because the word order and intonation are fixed (Darwin & Gray, 1999), any deviation from which leads to syntactic incoherence. This unique characteristic within multi-word phenomena, a 'bridge' between lexis and syntax, led Gass and Selinker (2001) to argue for the importance of phrasal verbs in second language acquisition. Another reason the present researcher chose to use phrasal verbs in the OWLS is that there is evidence to suggest that despite the frequency and importance of phrasal verbs, L2 learners of English find them difficult to acquire (Moon, 1997; Darwin & Gray, 1999; Celce-Murcia & Larsen Freeman, 1999). This may be because many phrasal verbs have multiple meanings: Gardner and Davies (2007) investigated the British National Corpus (BNC) to find that the most frequent phrasal verbs have an average of 5.6 possible distinct meanings each. As a result, many L2 learners of English (even those whose L1 also contains phrasal verbs) choose to circumvent the problem by not using phrasal verbs at all (Liao & Fukuya, 2004; Hulstijn & Marchena, 1989).

The three phrasal verbs chosen for inclusion in the OWLS were:

- Turn out (come)
- Take over (assume control)
- Pick up (collect)

These phrasal verbs were chosen with reference to the work of Gardner and Davies (2007), Shin and Nation (2008) and Martinez and Schmitt (2012), who used analysis of corpora to investigate the frequency of phrasal verbs, collocations, and phrasal expressions respectively. The work of Gardner and Davies (2007) resulted in a list of the top 100 most frequent phrasal verb lemmas from the BNC. They used a very simple definition of the term phrasal verb (which made no acknowledgement of semantic transparency): a lexical

verb followed by an adverbial particle that is either contiguous or noncontiguous to that verb. Two computer programmes were used to search the corpus: one to find occurrences of a lexical verb followed by an adverbial particle, the other to count the different meaning senses of the phrasal verbs found. Although in the resulting list we are told how many meaning senses were recorded for each phrasal verb, we are not told the frequency of each meaning sense, and therefore it is not clear which meaning is most frequent, or indeed whether the least frequent meaning sense of one phrasal verb is more frequent than the most frequent meaning sense of a lower-ranked phrasal verb.

The list provided by Shin and Nation (2008) ranked the 100 most frequent collocations from the spoken section of the BNC (10 million words). They operationalised the construct 'collocation' by using a computer programme to search for pivot words (word types of either nouns, verbs, adjectives or adverbs), and applied a number of cumulative criteria: All pivot words had to correspond to the 1,000 most frequent words from Leech, Rayson and Wilson's (2001) spoken word frequency list so that the collocations found would be likely to build on elementary level learners' existing lexical knowledge. All collocations had to occur more than 30 times within the ten million running BNC words, and could not violate an "immediate constituent boundary" (2008:342); that is to say, sentences were divided into phrases, and then these phrases subdivided as far as a minimum of two words. Each separate meaning was given its own count. The resulting list of 100 collocations includes eight phrasal verbs yet the most frequent meaning of each collocation is not provided, so we do not know which semantic variation of the phrasal verbs included is being referred to.

Investigating the frequency of 'phrasal expressions', Martinez and Schmitt (2012) compiled a 505-strong list which they divided, similarly to Nation's (1990) levels test, into five frequency levels; 1,000 – 5,000. They define their term as "a…sequence of two or

more co-occurring…words with a cohesive meaning or function that is not easily discernible by decoding the individual words alone" (2012:6) and clearly therefore, their construct encompasses phrasal verbs. The researchers specified three core criteria: Each expression must function as a morpheme equivalent (i.e. it is processed in the same way as a single-word unit), or be semantically opaque, or be "deceptively transparent" (2012:10). Additionally they used three auxiliary criteria: Expressions needed to have a single-item equivalent, or they could potentially be negatively influenced by a learner's L1, or their meaning/opacity could be altered as a result of the syntax of the expression. In the resulting list, 118 of the 505 expressions are phrasal verbs (23%): three phrasal verbs occur in the first 1,000 expressions, 33 in the second, 25 in the third, 31 in the fourth, and 26 in the fifth. Improving on Shin and Nation's (2008) work, Martinez and Schmitt usefully included a sentence using each expression in context so that the specific meaning was clear.

To mirror the decision taken in choosing the single-item target words, the present researcher opted to utilize phrasal verbs whose frequency suggested they are most useful for learners to acquire. The phrasal verbs from the three lists discussed above were divided into two groups; PV1 (Phrasal Verb 1) containing higher frequency phrasal verbs, and PV2 containing lower frequency phrasal verbs. In the case of Gardner and Davies (2007), who included only phrasal verbs in their list, the first most frequent 25 items went into PV1, and the second 25 most frequent into PV2. Shin and Nation's (2008) list covered a broader term ('collocations') and therefore there were far less phrasal verbs within their 25 most frequent collocations. As a result, all phrasal verbs from their list that occurred in the first 50 most frequent were placed in PV1, and those included in the second most frequent 50 were placed into PV2. In the case of Martinez and Schmitt (2012), as so few phrasal verbs occurred in their 1k frequency level, all phrasal verbs from the 1k and 2k levels were

placed in PV1, and all phrasal verbs from the 3k level were placed in PV2. To broaden the available options to choose from within these two frequency groups, the definition of phrasal verbs as used by Gardner and Davies (2007) was used. The phrasal verbs shown in table 3.2 are those that occurred in two or more of the three lists in PV1 and PV2.

Table 3.2: The most frequent phrasal verbs

PV1	PV2
Come on	Go off
Come in	Go in(to)
Find out	Get on
Come back	Turn out
Go on	
Pick up	
Take over	

Of these phrasal verbs, the three chosen for inclusion in the OWLS were selected because the researcher felt (based on extensive EFL teaching experience) that these three were least likely to be familiar to pre-intermediate level learners and above, and because they were amongst the most semantically opaque (and therefore potentially more difficult to acquire).

The three most recently added simulations were devised so that the target word was presented to the participant differently in each. That is, in simulation 6 (‘turn out’), the stimulus is a spoken monologic statement, indirectly addressed to the participant. In simulation 7 (‘take over’), the stimulus is spoken, is in question form, forms part of a dialogue and is directed at the participant. Simulation 8 (‘pick up’) presents a monologic written stimulus in a statement. The reason for presenting the stimuli in these ways was so that, at the data analysis stage, comparisons could be made between participants’ strategic responses to the different stimulus presentation forms and word types. The presentation forms of all eight OWLS simulations are shown in table 3.3.

Table 3.3: Stimulus presentation forms in the OWLS

	Type	Statement	Question	Dialogic	Monologic	Spoken	Written	Direct	Indirect
1	adj.		✔	✔		✔		✔	
2	n.	✔		✔		✔			✔
3	vb.	✔			✔		✔	✔	
4	n.	✔		✔		✔			✔
5	adj.	✔			✔	✔			✔
6	phr.vb.	✔			✔	✔			✔
7	phr.vb.		✔	✔		✔		✔	
8	phr.vb.	✔		✔	✔		✔	✔	

The contexts in which to present the stimuli in simulations 6, 7 and 8 were chosen based on the scenarios most identified with in the 2014 study, as shown in table 3.4.

Table 3.4: The six most identified with scenarios Briggs (2014)

	Scenario	**Mean**	**S.D.**
1	I listen to other people's conversations in English outside of class.	1.43	1.03
2	I use English for short exchanges with the people I live with (e.g. greetings / "Please pass the salt" / "I'm leaving").	1.70	1.15
3	I listen to English songs outside of class.	1.73	1.07
4	I speak in English to service personnel (e.g. shop assistant, bank clerk).	1.77	1.06
5	I listen to English TV/spoken radio programmes/podcasts outside of class.	1.83	1.33
6	I read timetables, announcements, posters, menus etc. outside of class.	1.91	1.08

The context of simulation 6 (watching the news) fitted scenario five, the context of simulation 7 (conversing at home with a cohabitant) fitted scenario two, and the context of simulation 8 (reading a billboard advertisement) fitted scenario six.

The three new simulations were created in PowerPoint as per the simulations used in the 2014 study, and piloted in the same way. A summary of the scenarios simulated in the version of the OWLS which was used in the present study, including the target word used in each simulation, is shown in table 3.5.

Table 3.5: Scenarios simulated in the OWLS in the present study

Simulation	Questionnaire item number	Scenario	Target word
1	6	Use English for short exchanges with cohabitants	fellow
2	22	Listen to English DVDs/movies	nuisance
3	20	Read timetables, announcements, posters, menus etc.	urge
4	3	Speak in English to service personnel	bargain
5	25	Listen to English songs	vain
6	8	Listen to English TV/radio/podcasts	turn out
7	6	Use English for short exchanges with cohabitants	take over
8	20	Read timetables, announcements, posters, menus etc.	pick up

3.5.4 INTERVIEW SCHEDULE

The interview schedule (Appendix G) was designed to encourage participants to voice what they would do in their heads in response to each simulation, not only or primarily their physical actions. The schedule was designed to be cyclical as a means of eliciting participants' deployment of more than one vocabulary acquisition strategy in response to each individual simulation, and whether that group of strategies constituted a cluster or a chain.

The interview schedule was piloted in conjunction with the OWLS prior to the 2014 study. The n=8 pilot participants were asked if they had had any response to the simulations which had not been elicited by the schedule, and if so, to elaborate on said responses. All 8

pilot participants reported that the schedule, applied cyclically, was exhaustive in terms of eliciting their actions and intentions in response to the simulations. This proved to be the case in Briggs (2014), and as such the researcher used the interview schedule in the present study unchanged from the 2014 study.

3.5.5 STRATEGIES TAXONOMY

The operationalisation in this study of the term 'vocabulary-related strategy' is presented, followed by an account of the development of the strategies taxonomy.

Dörnyei and Skehan (2003) propose that strategies must be defined as either neurological *or* cognitive *or* behavioural. In this study, the term 'vocabulary-related strategy' is firmly operationalised as cognitive; that is, any action taken by the learner behind which their self-reported *intention* is to promote vocabulary comprehension or acquisition. An observable behavioural act alone cannot constitute a strategy; to be deemed strategic, motor processes must be explained by an acquisition-driven, cognitive process. E.g., in conversation with their host family, a study abroad learner writes down a new word they hear. When the researcher asks why they wrote that word, the learner replies that their teacher told them to make a note of new vocabulary. In this instance, the learner has *not* employed a strategy because the intention behind their behavioural act was primarily to comply with their teacher's instructions, not to acquire or better to comprehend the word they wrote down.

Similarly with verbal action (i.e. using a new lexical item in speech), the learner is not employing a vocabulary-related strategy unless their intention at the point of speaking is to acquire that item. E.g., in conversation 1 below, the learner is not acting strategically, whereas in conversation 2, repetition of the word in order to picture its spelling could either act as a memorisation strategy, or as determination (e.g. to identify affixes/roots to

determine part of speech). Of course, in this case, a researcher would need to probe why picturing the spelling is helpful to the learner to identify which strategy is being employed.

Conversation 1

Host Mother:	John's at the theme park. He says he's already been on five rollercoasters.
Learner:	What?
Researcher:	Why did you say 'what'?
Learner:	I didn't hear what she said.

Conversation 2

Host Mother:	John's at the theme park. He says he's already been on five rollercoasters.
Learner:	What?
Researcher:	Why did you say 'what'?
Learner:	I didn't understand the last word.
Researcher:	How does it help you understand if you hear it again?
Learner:	I can try to imagine how it's spelt.

In the case of unobservable behaviour, if a learner engages in cognitive processes that might serve to enhance acquisition, yet their intention was not specifically to do so, this action also cannot be classified as strategic. E.g., the researcher observes a learner in conversation with a native English speaker as per conversation 3.

Conversation 3

Host Mother:	John's at the theme park. He says he's already been on five rollercoasters.
Learner:	*Pause*
Researcher:	What were you thinking then?
Learner:	I repeated the word in my head.
Researcher:	Why did you do that?
Learner:	I like the way it sounds.

If reported as a stand-alone cognitive act, this mental behaviour could be deemed a memorisation strategy. However, with the added lens of the learner's intention, we see that neither memorisation (and nor, therefore, acquisition) was the driving factor.

In order to identify strategic behaviour within the confines of the term as operationalised within in this study, 2 questions must always be answered:

1. What did the learner do? and
2. Why did they do it?

In this way, the 'Outside-Inside Problem' (Stevick, 1990) is addressed because the relationship between action and mental constructs is established.

The taxonomy of vocabulary-related strategies used in this study (Appendix H) was originally developed for a study undertaken by the researcher in 2010 and again used in Briggs (2014). It is based on those of Schmitt (1997), Nation (1990), Gu and Johnson (1996) and Oxford (1990). However, none of these existing taxonomies is exhaustive, and none were developed to specifically address the out-of-class behaviour of L2 learners. All of the taxonomies include strategies which can only be applied in classroom settings, e.g.:

"Ask teacher for a sentence including the new word" (Schmitt, 1997:207)

Furthermore, many of the strategies included in these taxonomies are unspecific about the mental action of learner. E.g.

"Consulting a reference source" (Nation, 1990:220)

fails to specify what kind of reference source is being consulted (e.g. an interlocutor, a search engine etc.). Neither does it stipulate the intention behind the action (i.e. what information does the learner want to find from this reference source?). It may be that a

learner intends only to comprehend, and not to acquire, the target word/phrase. Additionally, many strategies listed do not state whether an action is mental (and therefore unobservable) or physical. Thus, the taxonomy used in the present study has been developed to make explicit learner intention, and to relate only to strategies available in informal contact.

The taxonomy was piloted prior to the 2010 study using two L1 Malay speakers of advanced proficiency in English. They were individually interviewed and asked to discuss what they do to acquire new vocabulary encountered outside of the classroom. The interviews were then transcribed and the data analysed to ascertain whether the taxonomy accounted for all of the behaviours they reported. As this proved to be the case, no changes were made to the taxonomy as a result of piloting.

Although in the 2010 study the strategies taxonomy went some way to addressing vocabulary-related strategies in a more specific manner than pre-existing efforts, it became apparent in Briggs (2010, 2014) that it was not exhaustive. E.g., under the category 'Ask', participants reported strategies which involved asking others, but with an intent which was not covered by the taxonomy (i.e. repetition rather than working out meaning). Similarly, the taxonomy did not cover strategies that aimed to expose the learner to new vocabulary. To address these issues in the present study, the taxonomy was updated to include asking for repetition and creating opportunities for exposure to new lexis. Similarly, the present version of the taxonomy has been added to in order to encompass strategic use of modern technology, e.g. using search engines to find specific information about lexical items.

The taxonomy in the present study comprised 47 strategies divided into five categories: Planning, Individual determination of meaning, Interactive determination of meaning, Memorisation, and Metacognition. A distinction was made between the categories of

Planning and Metacognition as a means of differentiating between metacogntive strategies that are stand-alone (for example, deciding whether or not to engage with a new word/phrase) and metacognitive strategies which involve the consideration of other strategies (for example, evaluation of success of strategic behaviour against set goals).

3.6 PROCEDURE

The study proceeded in three stages in order to complement the mixed methods data collection employed in the study design. Figure 3.2 illustrates the timings of the data collection process stage by stage. Data collection began in May 2012 and concluded in December 2013.

Figure 3.2: Data collection timings

VT = vocabulary test; Q = questionnaire; OWLS = simulation interview

3.6.1 RECRUITMENT

The researcher contacted by email and telephone a large number of private language institutions in Oxford and London. The Director of Studies, Academic/Courses Manager or Owner of those institutions who responded positively to this initial contact were visited in person by the researcher in order to discuss in person what participation would involve at both institution and student level. If after that initial meeting an institution was willing to participate, an initial recruitment date was set. In all cases, the initial and subsequent

recruitment dates were set for two to three days after a new intake of students was expected at the institution. It was agreed with all participating institutions that before each recruitment day, the study information letter (Appendix A) would be distributed to new students; either distributed by hand as part of an introduction to the study given orally at a welcome meeting for new students on their first day at the institution, or included in the institution 'welcome pack' given to each new student on arrival. The participating private language institutions also agreed in the initial meeting to provide the researcher with information in advance of the number of new students expected from week to week, the arrival and (where known) departure dates of these students, and their proficiency level as self-reported prior to arrival in the UK.

The procedure on each recruitment day was usually as follows: The researcher was informed by the institution of which classrooms contained new students (of pre-intermediate proficiency level and above - as determined by placement testing carried out by the institution - who were studying at the institution for six weeks or more) during a specific lesson in the day (most often the last lesson before the lunch break). The researcher would then enter the teachers' room at mid-morning break to introduce herself or reacquaint the teachers to her, and to ask each teacher if there was a specific time during that lesson slot which would be most convenient in which to interrupt and enter the class. The researcher then interrupted each class for about two minutes at the agreed time, introduced herself to the students and in particular the new students, explained in brief the purpose of the study and what participation would involve and then, crucially, told the new students where to go later that day for the recruitment meeting, in which they could sign up to and/or ask questions about the study. The recruitment meeting always took place that same day and was held in a vacant classroom at the institution, usually from half an hour into the lunch break and for the remainder of the afternoon.

Occasionally the institution arranged a meeting, usually in the afternoon of the students' second day, to which all new students who matched the sampling frame were asked by the institution to attend and in which the researcher was able to present the details of the study and sign up those willing to participate.

The vast majority of students who attended a recruitment meeting signed a consent form at the meeting and completed the time 1 vocabulary test that same day. Those who did not complete the test that day signed the consent form and exchanged contact details with the researcher, and a subsequent testing day was arranged at the earliest possible convenience within two weeks of their arrival.

3.6.2 VOCABULARY TESTING

The time 1 vocabulary test was administered as soon as possible after recruitment to the study, and within two weeks of a participant's arrival to the UK. The majority of the participants took the time 1 test on the same day as they signed the consent form, that is, they took the test directly after the recruitment meeting at their school.

The time 1 test was administered on-site at the learners' particular language institution, in a classroom and under test conditions. Participants were told that they would take another vocabulary test at the end of their SAE, but (to negate any practice effect) not that the second test would contain the same items. An example of the question types in both part one (receptive knowledge) and part two (productive knowledge) were shown to the participants and understanding was checked using a series of pre-devised checking questions. These were:

Part One: How many vocabulary items are there in each question? (6). How many definitions? (3). How many numbers do you write next to each definition? (1). What is the

relationship between the number and the definition? (They have the same meaning). How many vocabulary items will you NOT use in your answer? (3).

Part Two: Where do you write your answer? (In the space/on the line). How many words do you need to write? (1).

The instructions were checked in this way to ensure the participants had fully understood the task at hand (because the rubric was written in the L2). Additionally, comprehension-checking questions such as these were deemed to be familiar to the participants because EFL teaching qualifications (e.g. Cambridge ESOL Examinations' CELTA, DELTA) require trainees to check instructions in this manner.

The researcher advised participants that there was only one correct answer for each question. For part two, they were told not to be overly concerned with spelling; that if the researcher was able to understand the word and it was the correct answer, they would earn a mark regardless of spelling errors. The participants were given forty minutes in total to complete both parts of the test, and were independent in their decision about how much time within those forty minutes to spend on each part, and on the test in total.

The administration of the time 2 vocabulary test matched exactly that of time 1, and took place no more than two weeks prior to the end of the learner's SAE, or to the end of their participation in the study.

3.6.3 QUESTIONNAIRE ADMINISTRATION

Participants completed the questionnaire no sooner than two weeks after their arrival in the UK, in their language institutions and in the presence of the researcher. A period of at least 2 weeks before completion of the questionnaire was decided upon because the researcher's previous contact with study abroad learners revealed that they are sometimes placed in

temporary accommodation for their first week in England, before being moved to the host family/dorm where they will stay for the remainder of their SAE. Additionally, a fortnight's stay allowed for enough time to have passed for participants to experience the scenarios about which they were questioned in this instrument but it was also a short enough period that those staying only for a short time (e.g. group 1) 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.

3.6.4 SEMI-STRUCTURED INTERVIEWS

With the permission of the Director of Studies/manager/owner at the respective language institutions, all interviews took place in a classroom, self-access study room or library at the institutions. The interviews took place at least two weeks after the completion of the questionnaire. Two weeks was arrived at because some study abroad learners stay in the UK for a little as six weeks (particularly in the summer holidays), so to ensure their complete participation a short lead-time between these data collection stages was necessary. Where participants had a longer SAE (groups 2 and 3), more time passed between completion of the questionnaire and taking part in the interview. However, the researcher aimed as far as possible to keep the amount of time between data collection stages very similar for participants in the same group.

The researcher conducted the interviews equipped with a laptop computer on which to play the OWLS, an audio recorder to record the interview, a hardcopy of the interview schedule (not shown to the participant), and a pen and paper to make notes if necessary (e.g. if the audio recorder failed to operate). The interview began with the researcher gaining verbal consent from the participant to audio-record the interview. The researcher then told the participant that they would be shown some situations on the laptop, and that

each situation comprised two slides. The researcher explained that after the second slide of each situation had finished playing, the participant would be asked some questions about their reaction to the situation.

The researcher emphasised to the participant that the interview was not a test, that there were no right or wrong answers, and that if at any time they wanted to ask a question of the researcher, they should feel free to do so. The researcher then played the OWLS, and after the audio track on each of the stimulus slides had finished, proceeded to ask the interview schedule questions. At the end of the interview, the researcher asked the participant if there was anything they would like to add, thanked them for their participation, and turned off the audio recorder. On average, an OWLS interview lasted for approximately 30 minutes.

3.6.5 PROCEDURAL LIMITATIONS

Particularly for group 1 (6-10 weeks), there may have been an effect of the earlier stages of data collection on the latter stages. For example, the shortest period of time between the time 1 and time 2 vocabulary tests was six weeks. Although participants were informed that they would be tested again at the end of their SAE, they were not told that the same items would be tested, and to negate further any practice effect the items of the time 2 test were presented in a different order. However, the knowledge that the focus of the study was that of vocabulary gain may have impacted participants' behaviour in terms of the new lexis they encountered during the study.

Part 1 of the questionnaire (identification with informal L2 contact) has the potential to have suggested L2 contact scenarios to the participants which they had not previously considered. If they had sought out these opportunities as a result of completing the questionnaire, there may have been an effect on their responses to the OWLS instrument

(because they had more opportunity to develop strategic responses to these scenarios) and potentially on their performance on the time 2 vocabulary test (because the heightened exposure to the L2 had positively impacted vocabulary gain). Similarly, part 2 of the questionnaire may have suggested VLS to the participants that they would otherwise not have used, and this in turn may have impacted on their responses to the OWLS and to the time 2 vocabulary test.

The OWLS interview could potentially have 'trained' participants in terms of how they responded to unfamiliar vocabulary in informal L2 contact during the remainder of their SAE. That is, by asking them to report in highly specific terms on their cognitive behaviour in the scenarios simulated, participants may have gained a heightened awareness of what strategies they were using, and how effective they deemed that behaviour. This could in turn have led to their modifying the strategies they used in these (and indeed other) scenarios, which could have impacted their vocabulary gain and thus their time 2 vocabulary test results.

Although the potential for interaction between the data collection stages and instruments used in this study presents limitations to the design and findings, the limitations refer only to the study. That is, greater awareness of available informal L2 contact scenarios and vocabulary-related strategies, and the opportunity for self-evaluation in terms of strategic behaviour, are both wholly positive factors for the participants. As no evidence yet exists to show a relationship between the three main variables in this study (informal contact, strategy use and vocabulary gain), the present researcher viewed the limitations here discussed as a worthwhile 'trade-off'; necessary evils to answer the research questions posed.

3.7 DATA ANALYSIS

The analysis strategies will now be discussed with reference to the research questions.

3.7.1 RESEARCH QUESTION 1

1. Is there a relationship between identification with informal L2 contact scenarios and vocabulary gain in a study abroad context?

    1a. Do location of SAE and length of stay play a role in this relationship?

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 in that they were marked by the researcher and by an inter-rater, who was an experienced English language teacher and teacher trainer.

The mean scores of the items from part 1 of the questionnaire (informal L2 contact) were calculated to determine the order in which the scenarios were most identified with by the sample. Exploratory factor analysis was performed using this data to determine whether the same underlying factors as in Briggs (2014) existed in the data from the present study. Subsequently, correlation analyses were used to explore whether or not there was an association between vocabulary gain and the factor-reduced out-of-class contact.

The continuous variable from the identification with informal L2 scenarios measure (which was the sum of scores in response to the 28 informal contact scenarios in part 1 of the questionnaire) was taken as the independent variable of primary interest in a series of hierarchical multiple regression analyses. In this continuous variable participants could score a minimum of 28 (if they had allocated '(1) This is very true of me' to every scenario) and a maximum of 140 (if they had allocated '(5) This is not at all true of me' to every scenario). In terms of the other two independent variables, the sample was divided

for the regression into a three-level ordinal variable based on length of stay: group 1 (6-10 weeks), group 2 (11-15 weeks) and group 3 (16-20 weeks) and into a binary categorical variable based on the location of the SAE (Oxford or London). Regression analysis was conducted using these three independent variable taking receptive, productive and then overall vocabulary gain as the dependent variable respectively.

3.7.2 RESEARCH QUESTION 2

2. What vocabulary-related strategies do learners of English employ in informal L2 contact in a study abroad context?

As per RQ1, to answer RQ2 the data pertaining to strategic behaviour from Part 2 of the questionnaire was subjected to factor analysis in order to reveal the underlying structure in the data. The semi-structured interviews were transcribed verbatim and analysed with reference to the strategies taxonomy. For each simulation, mentions of/allusions to strategies that were included in the taxonomy were quantified. If a strategy was mentioned which had not been included, it was added to the taxonomy and quantified in the same way. Where combinations of strategies were referred to, clusters and chains of strategies were also quantified. With each count of a strategy or combination of strategies, whether it was high or low inference (i.e. the clarity or ambiguity of the reported mental action/intention) was also noted. An inter-rater (an experienced teacher of L2 English) was used to enhance reliability by providing agreement on which reports were high and low inference, and deciphering the high inference reported strategies. A percentage of high/low inference codings was calculated.

3.7.3 RESEARCH QUESTION 3

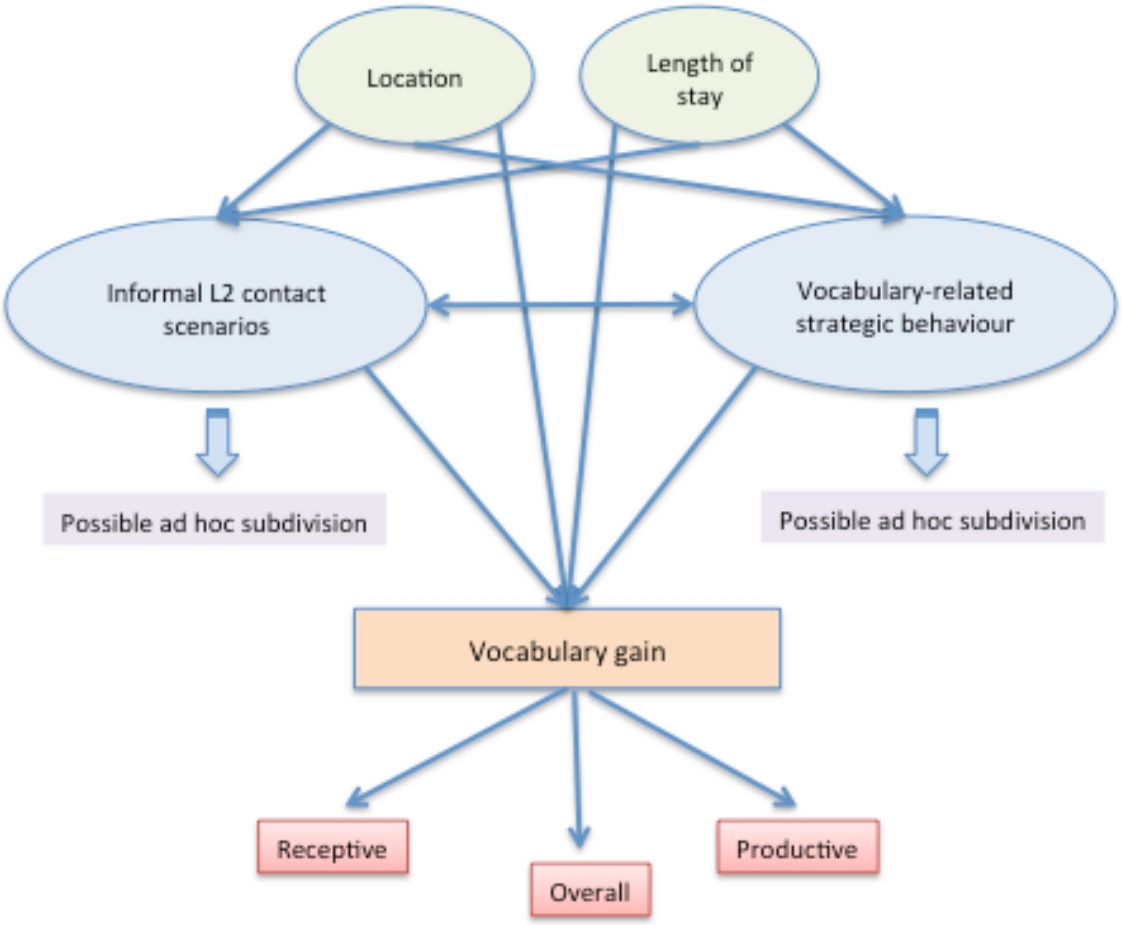
3. Is there a relationship between vocabulary-related strategy use in informal L2 contact and vocabulary gain in a study abroad context?

    3a. Do location of SAE and length of stay play a role in this relationship?

The approach to answering RQ3 was similar to that of research question 1 in that the data was manipulated to determine whether or not a correlation existed between the two variables addressed, and then multiple regression employed to investigate which factors of strategic behaviour (as defined by factor analysis) most strongly predicted receptive, productive and overall vocabulary gain.

Figure 3.3 shows all the variables analysed for the three research questions.

Figure 3.3: Study variables



3.8 ETHICAL CONSIDERATIONS

This project was reviewed by and received ethics clearance through the University of Oxford Central University Research Ethics Committee (CUREC). Evidence of CUREC approval for the study is provided in Appendix B. Ethical considerations are further approached in two ways: Firstly, a discussion of how the present study encompasses and meets the six recommendations for good practice in applied linguistics student projects set forth by The British Association for Applied Linguistics (BAAL; 2000). Secondly, a more general discussion is presented of the moral-philosophical considerations pertaining to second language acquisition research.

The recommendations made by the BAAL (2000) document pertain to the responsibility of the researcher to their participants. In the present study, participants' rights were fully respected: All potential participants received a study information letter (Appendix A) prior to asking for their consent to participate, and were given the opportunity to speak in person to the researcher or to contact the researcher by e-mail or telephone to clarify any points made therein. The letter set out the aims of the study, what participation would involve in terms of time commitments and information sought, and how participant confidentiality would be upheld. Additionally, the letter advised participants of their right to withdraw from the study at any time, without the need for explanation, as a means of highlighting their freedom from coercion. Participants gave their written consent at the point of recruitment (Appendix C). This piece of research did not involve deception or covert practices of any kind.

There was a wide variety in the sample of nationality, culture, age and gender, and the researcher was careful to consider and respect these diversities. For example, the researcher was careful to be entirely flexible to participants' preferences in terms of when

to meet for data collection and testing, so as not to interfere with, for example, religious practices (e.g. congregational prayer, Ramadan) or personal circumstances that may have impacted on their availability/ability to engage in the data collection procedure. Additionally the researcher sought to minimize any potential disruption to the lives and environments of the participants by conducting data collection at their language institutions; an environment with which they were comfortable and familiar, and to which no extra travel constraints were imposed.

The BAAL (2000) document also highlights the importance of confidentiality and anonymity of participants in applied linguistics research. To address this issue in the present study, the study information letter (Appendix A) informed potential participants that their real names would not be used and that they would not be identifiable in any way. All information provided was kept on a secure, password-controlled computer to which only the researcher and her supervisor had access. The information given was only used for the purpose of writing about the study. If direct quotations were used, these were sent in advance by e-mail to the participant who provided them, and their consent sought before use.

In terms of broader ethical considerations, the researcher was aware that the population under investigation in this piece of research was not representative of all second language learners. That is, that social and economic bias disallowed a large proportion of learners of English the opportunity to study and live within the target language community. It is hoped that this ethical burden was mitigated somewhat through the rigorous application of the highest standards of research conduct: If a study finds in favour of certain types of informal contact and/or the employment of certain strategies/groups of strategies on L2 vocabulary gain in a study abroad context, and if these findings stand up to critique and are deemed reliable, then what is learned from the study has the potential to be applied to

informal contact and/or strategy use in other contexts and thus to a wider population of language learners.

CHAPTER IV

FINDINGS PART I

INFORMAL L2 CONTACT & VOCABULARY GAIN

The findings reported in this chapter pertain to the first research question that the study sought to address:

1. Is there a relationship between identification with informal L2 contact scenarios and vocabulary gain in a study abroad context?

   1a. Do location of SAE and length of stay play a role in this relationship?

4.1 VOCABULARY GAIN

In this section the findings related to the attributes of vocabulary gain, the dependent variable, are investigated and described. Vocabulary gain was measured at two time points: time 1 within a fortnight of a participant's arrival to the UK, and time 2 within a fortnight of their departure.

4.1.1 INTER-RATER RELIABILITY

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 the inter-rater. 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). According to Landis and Koch (1977), a value higher than K=.8 indicates a very high strength of agreement.

4.1.2 DESCRIPTIVES AND NORMALITY

Here follows the discussion of the descriptive statistics and normality of distribution of the following vocabulary variables respectively: time 1 and 2 raw scores, time 1 and 2 percentage scores, and percentage gain.

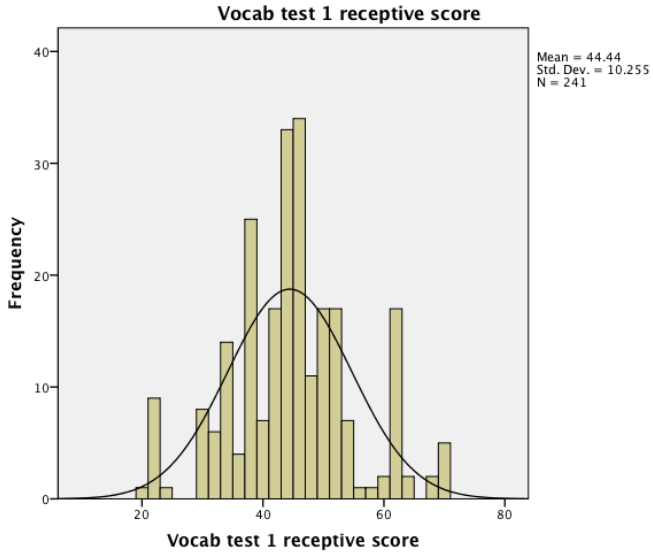
4.1.2.1 Time 1 raw scores

N=241 participants completed the time 1 vocabulary test. The data scored for receptive vocabulary knowledge out of a possible maximum score of 72, and for productive vocabulary knowledge out of a possible maximum score of 40.

4.1.2.1a *Time 1 receptive raw scores*

The scores on the time 1 receptive test ranged from 22 to 70 (M=44.44, SD=10.26). This variable was normally distributed, with a skewness of .154 (SE=.157) and a kurtosis of .268 (SE=.312). An inspection of the histogram for this variable supported this assertion, with the approximate symmetry demonstrating the skewness value and the high peaking of the central scores illustrating the more positive kurtosis value. The assumption of normality was further supported by reference to the 5% trimmed mean (44.43), which indicated that any outlaying cases were not having an impact upon the mean score.

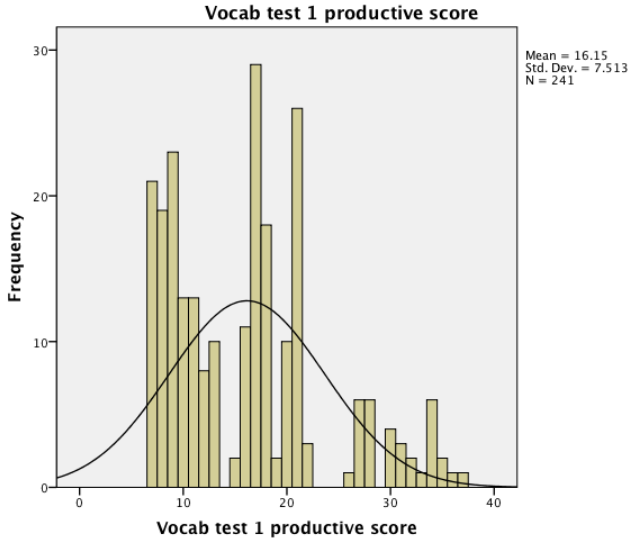
Figure 4.1: Distribution of time 1 receptive raw scores



4.1.2.1b *Time 1 productive raw scores*

The scores on the time 1 productive test ranged from 7 to 37 (M=16.15, SD=7.51). This variable was non-normally distributed due to a moderate skewness value of .795 (SE=.157). Kurtosis was normal at -.006 (SE=.312). The sig. value of the Kolmogorov-Smirnov statistic was .000, further suggesting violation of the assumption of normality.

Figure 4.2: Distribution of time 1 productive raw scores



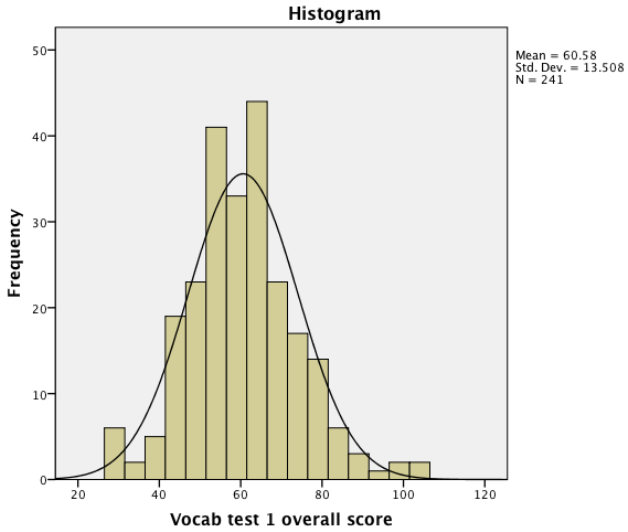
However, inspection of the boxplot revealed that SPSS found no outliers in this variable,

which may account for the similarity to the mean value of the 5% trimmed mean (15.64).

4.1.2.1c *Time 1 overall raw scores*

The maximum possible overall raw score for the vocabulary test was 112. The scores on this variable ranged from 29 to 106 (M=60.58, SD=13.51). Inspection of the histogram against the plot of normal distribution suggested that the data were largely normally distributed with a slightly positive kurtosis level.

Figure 4.3: Distribution of time 1 overall raw scores



The skewness value was normal at .388 (SE=.157) but, as indicated by the histogram, there was a positive kurtosis value at the level of .729 (SE=.312). The sig. value of the Kolmogorov-Smirnov statistic was significant at .008, indicating that the data were in fact non-normally distributed. However, the 5% trimmed mean value of 4 outlaying cases in this variable as identified by SPSS had little to no effect upon the mean, suggesting that this variable did not deviate significantly from normality.

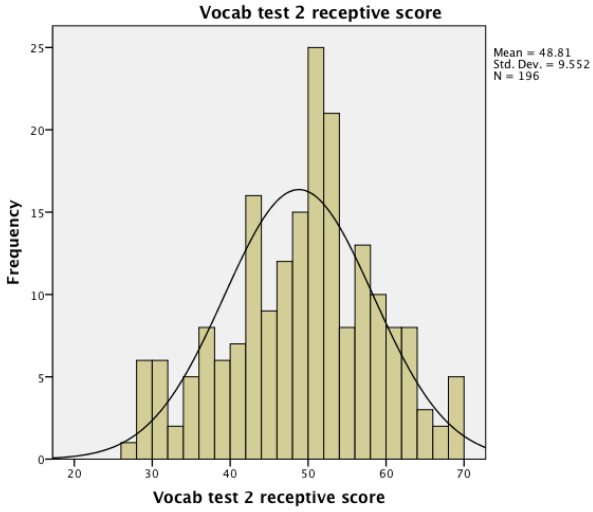
4.1.2.2 Time 2 raw scores

As a result of 24% participant mortality, there was a difference between the sample size at time 1 (n=241) and time 2 (n=196).

4.1.2.2a *Time 2 receptive raw scores*

The scores on the time 2 receptive vocabulary test ranged from 27 to 69 (M=48.81, SD=9.55) of a possible 72 marks. The skewness value was -.201 (SE=.174), suggesting that the data were approximately symmetric, and the kurtosis value was -.377 (SE=.346). Inspection of the histogram suggested near-normal distribution, an indication supported by the similarity of the 5% trimmed mean (48.89) to the mean.

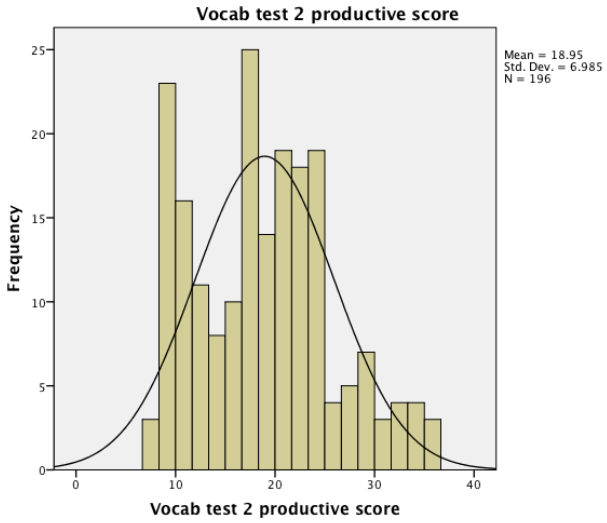
Figure 4.4: Distribution of time 2 receptive raw scores



4.1.2.2b *Time 2 productive raw scores*

Scores on the time 2 productive vocabulary test ranged from 7 to 36 (M=18.95, SD=6.99) of a possible 40 marks. The skewness value was .369 (SE=.174), illustrated by a skew to the left of the histogram, and the kurtosis value -.490 (SE=.346).

Figure 4.5: Distribution of time 2 productive raw scores

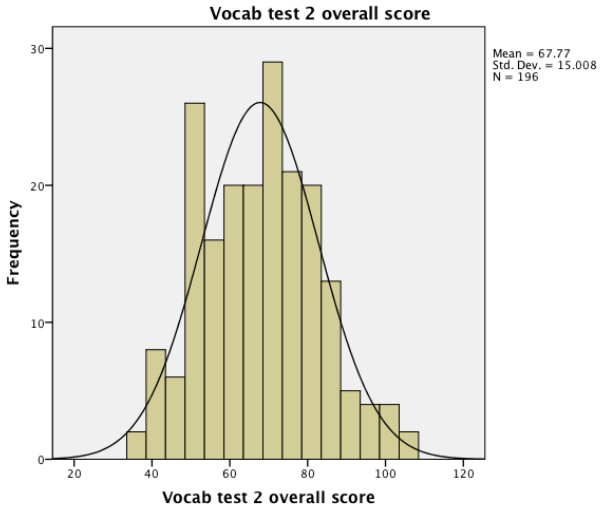


The sig. value of the Kolmogorov-Smirnov statistic was significant at *p*=.017, indicating that the data were in fact non-normally distributed. However, the 5% trimmed mean (18.68) had little to no effect upon the mean, suggesting that the variable did not deviate significantly from normality.

4.1.2.2c *Time 2 overall raw scores*

In terms of the receptive and productive time 2 vocabulary data combined, scores ranged from 36 to 106 of a possible 112 marks. The mean overall score was 67.77 (SD=15.01). Inspection of the histogram indicated a near-normal distribution of scores, with a very slight skew to the left of the graph. This indication was supported by a skewness value of .150 (SE=.174). The kurtosis value was -.383 (SE=.346), indicating a slightly platykurtic distribution.

Figure 4.6: Distribution of time 2 overall raw scores



The 5% trimmed mean of the overall time 2 variable was 67.58; the similarity of this value to the actual mean is further evidence in support of the observation that this variable is near-normally distributed.

4.1.2.3 Percentage scores

The means and standard deviations of the participants' total percentage scores on the time 1 and time 2 vocabulary tests (VT1 and VT2), in addition to their percentage scores on the receptive and productive parts of each test, were calculated. As previously stated, the sample size fell from n=241 at time 1 to n=196 at time 2 of the vocabulary test administration. Even so, due to the large sample size of the present study, tests to evaluate the skewness and kurtosis values may be over-sensitive (Tabachnick & Fidell, 2001) and thus an inspection of the histogram and boxplot for each of these test scores was conducted in order to assess the normality of the distribution of the scores.

Outliers were observed through inspection of the boxplots in VT1 overall and VT1 part 1 (receptive knowledge). After checking that these cases were in fact genuine scores, an inspection of the actual mean in comparison with the 5% trimmed mean in each of these 2

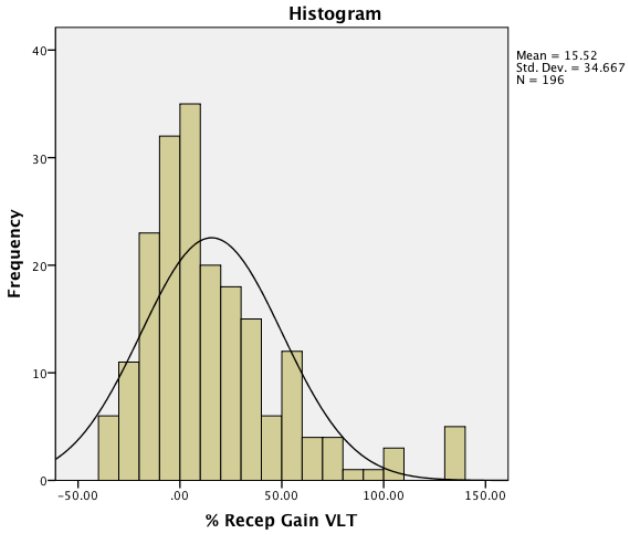
cases (respectively, 54.09 and 53.83; 61.72 and 61.71) supported the decision to retain these cases in the data file.

4.1.2.4 Percentage gain

4.1.2.4a *Receptive gain*

The receptive percentage gains between time 1 and 2 of the vocabulary test were calculated and this variable was analysed in terms of normality of distribution. There was a distinct difference between the mean and the 5% trimmed mean (M=15.52, SD=34.67; 5%TM=12.49). The Kolmogorov-Smirnov statistic was significant at p=.000 and the histogram showed a skew to the right, all of which indicated non-normal distribution.

Figure 4.7: Distribution of receptive percentage gain



There were no extreme outliers in this variable and therefore, in a bid to attain normal distribution without sacrificing valid, non-extreme outlying cases, a logarithmic transformation was carried out. Logarithm was considered appropriate in this case due to the positive skewness of the data. The logarithmically transformed data were no more normally distributed than pre-transformation. The receptive gain variable was then transformed using the square root and an inverse approach but neither had a more positive

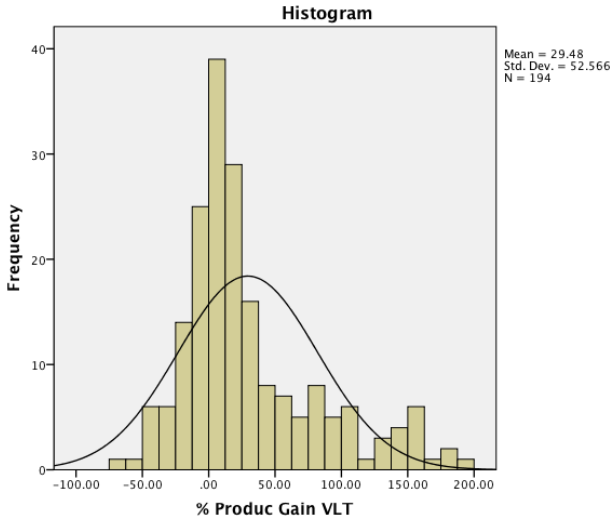
impact on the distribution of the receptive gain values.

As a result of the findings of these transformations, the original, non-transformed values were retained to represent the receptive vocabulary gain of the sample.

4.1.2.4b *Productive gain*

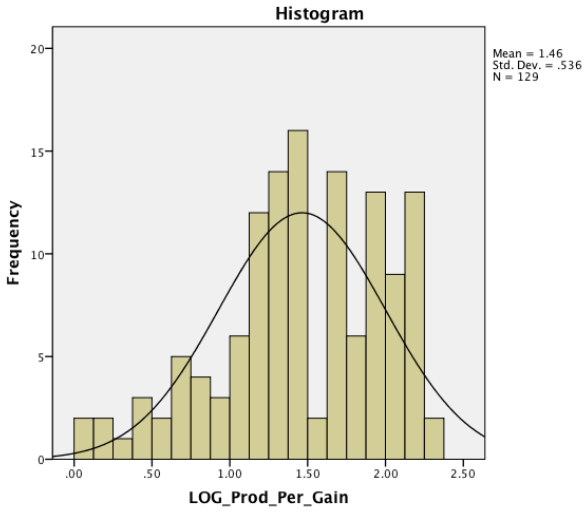
Inspection of the boxplot of cases for the productive vocabulary gain variable showed that there were eleven outliers, two of which were extreme. The outliers were having a significant impact upon the difference between the mean and the 5% trimmed mean (M=31.97, SD=4.13; 5%TM=27.30). These extreme cases were investigated and the researcher was able to confirm that they represented legitimate cases. These two extreme outliers were then removed from the dataset and the distribution re-examined, to find a smaller yet still not insignificant disparity (M=29.48, SD=52.57; 5%TM=26.13). Moreover, the Kolmogorov-Smirnov statistic was still significant at p=.000 and the histogram showed a distinct skew to the right-hand side even after the removal of the extreme outliers.

Figure 4.8: Distribution of productive gain



As before, this positively skewed variable was logarithmically transformed in the hope of achieving normal distribution. The logarithmically transformed productive gain variable was investigated to determine whether it conformed to normal distribution. The Kolmogorov-Smirnov statistic was found to be non-significant (p=.058) and there was very little difference between the mean and the 5% trimmed mean (M=1.46, SD=0.54; 5%TM=1.49). Inspection of the histogram against the plot of normal distribution showed a distinctly more normal distribution than pre-transformation and as such, the logarithmically transformed variables was retained to represent the sample's productive vocabulary gain.

Figure 4.9: Distribution of logarithmic productive gain



4.1.2.4c *Overall gain*

The overall (i.e. combining receptive and productive scores) percentage gains between time 1 and 2 of the vocabulary test were calculated and this variable was also analysed in terms of normality of distribution.

Table 4.1: Overall gain descriptives

Descriptive Statistics				
% Gain from VT1 to VT2 (Maximum possible gain of 10,000%)			Statistic	Std. Error
	Mean		15.7917	2.23052
	95% confidence interval for mean	Upper bound	11.3927	
		Lower bound	20.1908	
	5% trimmed mean		13.0397	
	Median		6.3550	
	Variance		975.147	
	Std. deviation		31.22735	
	Minimum		-29.08	
	Maximum		147.08	
	Range		176.16	
	Interquartile Range		33.27	
	Skewness		1.645	.174
	Kurtosis		3.632	.346

There was a substantial difference between the mean and the 5% trimmed mean (M=15.79, 5%TM=13.04) as a result of outlying cases in this variable. The skewness value also indicated non-symmetrical distribution. Moreover, the sig. value of the Kolmogorov-Smirnov statistic was .000, suggesting violation of the assumption of normality.

Table 4.2: Overall gain tests of normality

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
% Gain from VT1 to VT2	.147	196	.000	.866	196	.000

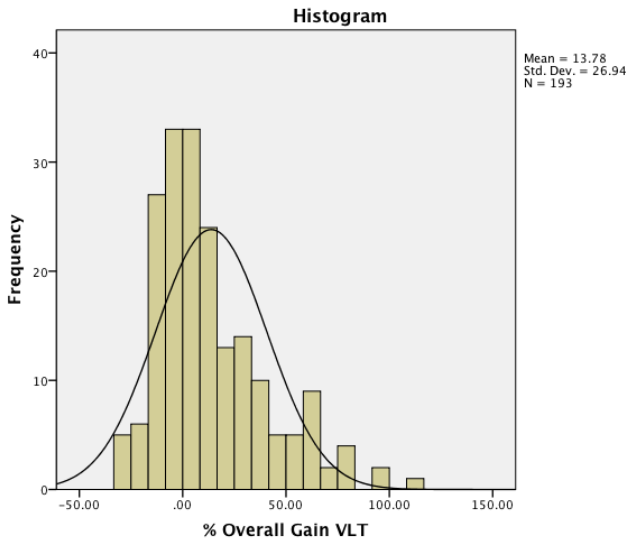
a. Lilliefors Significance Correction

Potential outliers could be observed to the right-most side of the histogram, and the boxplot confirmed this observation, displaying seven outliers of which three were deemed by SPSS to be extreme cases. The data pertaining to these three extreme outliers was re-examined to confirm that the raw scores and percentage scores had been calculated and entered into the dataset without error: These outliers represented legitimate cases.

There is disagreement among scholars regarding whether extreme yet legitimate outliers should be removed from a dataset (Osborne, 2002; Pirker, 2009). Judd and McClelland (1989) state that removal of outliers, regardless of legitimacy, enables the researcher to represent the population parameters in the most honest way possible. Conversely, Orr, Sackett and Dubois (1991) argue that retaining legitimate outliers enables a better representation of the population by the data. However, moving from the conceptual to the practical, there is robust evidence to suggest that the removal of outliers, particularly extreme cases, enhances accuracy and promotes a reduction in errors of interference when conducting inferential statistical analyses (Osborne & Overbay, 2004).

In the present case, the extreme scores in the vocabulary gain variable were removed and the variable re-inspected to determine the extent of the impact of the remaining (non-extreme) outliers on the 5% trimmed mean: this exploration found a far smaller disparity between the two figures (M=13.78, SD=26.94, 5%TM=12.06) and therefore the non-extreme cases were retained. However, the Kolomogorov-Smirnov statistic remained significant (p=.000) and the histogram displayed a remaining skew in the distribution of the values.

Figure 4.10: Distribution of overall gain



As a result of its non-normal distribution, the overall percentage gain variable was altered using logarithmic transformation. Logarithm was selected as the transformation method due to the positive skew of the distribution of this variable: in a logarithmic scale, the peak of the data is compressed whilst the tail is lengthened, and this process is therefore likely to more normally distribute the values in a positively skewed variable.

The logarithmically transformed overall vocabulary gain variable was investigated to determine whether it conformed to normal distribution, but the distribution of the transformed variable was less normal than that of the non-transformed variable. A square root and inverse transformation were then applied to the overall vocabulary gain variable data but did not more normally distribute the values in comparison with the original variable, nor with the logarithmically transformed variable.

In light of the failure of the data transformation techniques to more normally distribute this variable, the original data values were kept and the non-transformed variable retained.

4.1.2.5 Summary of descriptives and normality

The descriptive statistics and assessment of normality of the raw score variables indicated that the measures used to test vocabulary in this study were effective, appropriate, and well able to discriminate between the sample.

The explorations into the normality of the distribution of the vocabulary gain variables (receptive, productive and overall) demonstrated that in the case of the receptive and overall gain variables, normality of distribution was not possible to attain. However, as a result of logarithmic transformation, the productive vocabulary gain data were normally distributed.

In order to account for the non-normal distributions of the receptive and overall gain variables, on each occasion that a parametric statistical analysis was carried out using either or both of these variables, the equivalent non-parametric test was also conducted and reported in order to demonstrate the convergence or divergence of the findings of the two approaches.

4.1.3 COMPARISON OF MEANS

T-tests and analyses of variance (ANOVAs) were used to compare the means of divisions of the sample on the dependent variable, vocabulary gain.

4.1.3.1 *T*-tests

Two types of *t*-tests were conducted on the vocabulary test data: repeated measures and independent samples.

4.1.3.1a *Percentage scores at time 1 and 2*

A repeated measures/paired samples *t*-test was conducted on the time 1 and 2 percentage scores of the vocabulary test. In terms of receptive scores, there was a statistically significant gain from time 1 (M=61.49, SD=14.21) to time 2 (M=67.76, SD=13.25; $t(196)=-5.66$, $p=.000$) with a small effect size (Eta squared=0.14). With reference to productive vocabulary scores, the *t*-test found a significant increase between time 1 (M=40.25, SD=18.67) and time 2 (M=47.37, SD=17.48); $t(196)=-6.14$, $p=.000$) with a small effect size once more (Eta squared=0.16).

A repeated measures/paired samples *t*-test was also conducted to evaluate the impact of the SAE on the participants' mean percentage scores on the vocabulary tests overall (receptive scores + productive scores) between time 1 and time 2. There was a statistically significant gain in overall vocabulary scores from time 1 (M=53.90, SD=12.26) to time 2 (M=60.45,

SD=13.37); *t*(196)=-7.30, *p*=.000). The Eta squared statistic (0.21) indicated a small effect size.

These findings indicate that the sample improved their vocabulary knowledge between time 1 and time 2 and that a statistically significant improvement was made on their receptive scores considered independently, their productive scores considered independently, and on their overall vocabulary scores.

4.1.3.1b *Location and vocabulary gain*

An independent-samples *t*-test was conducted to reveal whether location impacted on receptive vocabulary gain. The significance level of Levene's test was *p*=.052 showing that the variances for the two location groups were the same. There was no significant difference in receptive vocabulary gain between participants in Oxford (M=16.36, SD=37.45) and participants in London (M=14.55, SD=31.33; *t*(194)=.363, *p*=.717). The results of the Mann-Whitney test of these variables also indicated no significant difference between the location groups [z=-.192, *p*=.848]. These findings suggest that location of SAE did not impact the receptive vocabulary gain of the participants in the sample.

The logarithmically transformed productive vocabulary gain variable was then used in an independent-samples *t*-test with location as the grouping variable. The significance level of Levene's test was *p*=.230 which indicated that the variances for the two location groups were the same. The *t*-test found no significant difference in vocabulary gain between participants in Oxford (M=1.34, SD=0.60) and participants in London (M=1.28, SD=0.50; *t*(113)=.558, *p*=.578). This finding indicates that there was no difference between the productive gain means of the Oxford and London groups.

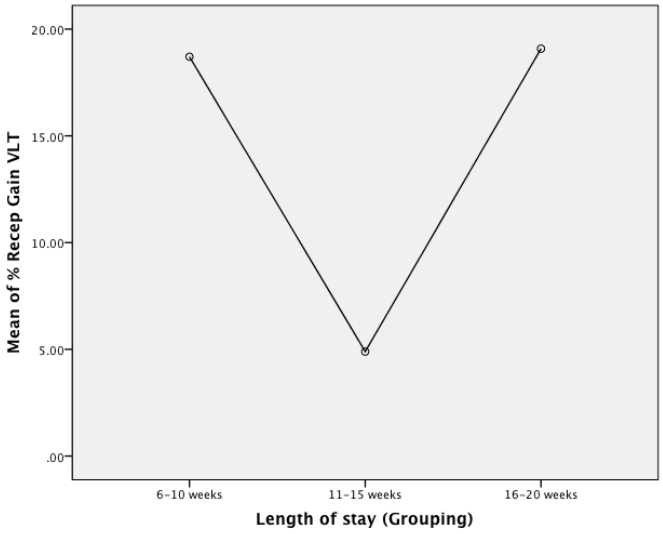
The overall percentage gain variable used as the dependent variable in an independent-samples *t*-test to determine whether there was a statistically significant difference in the overall vocabulary gain scores for the sample as grouped by location (Oxford and London). The significance level of Levene's test was p=.019 showing that the variances for the two location groups were not the same. There was no significant difference in vocabulary gain between participants in Oxford (M=14.89, SD=28.96) and participants in London (M=12.50, SD=24.47; t(191)=.622, p=.534). To ensure that the non-normal distribution of the overall gain variable was not adversely affecting this finding, a Mann-Whitney U test was also conducted, finding no significant difference between the means of the two groups [z=-1.76, p=.86]. These findings indicate that location of SAE did not impact on the overall vocabulary gain of the participants in this study.

4.1.3.2 ANOVAs

4.1.3.2a *Length of stay and vocabulary gain*

The receptive vocabulary percentage gain variable was taken as the continuous dependent variable in a one-way between-groups ANOVA to determine the impact of length of stay. The ANOVA revealed a statistically significant difference in receptive vocabulary gain for the three groups [F(2, 193)=3.05, p=.050]. The effect size was small (Eta squared=.03). The Tukey HSD post-hoc test indicated that, similarly to the findings of the ANOVA on overall gain, the mean 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 mean (M=18.71, SD=40.09) differed significantly neither from the medium nor the long stay group.

Figure 4.11: Means plot of receptive gain by length of stay



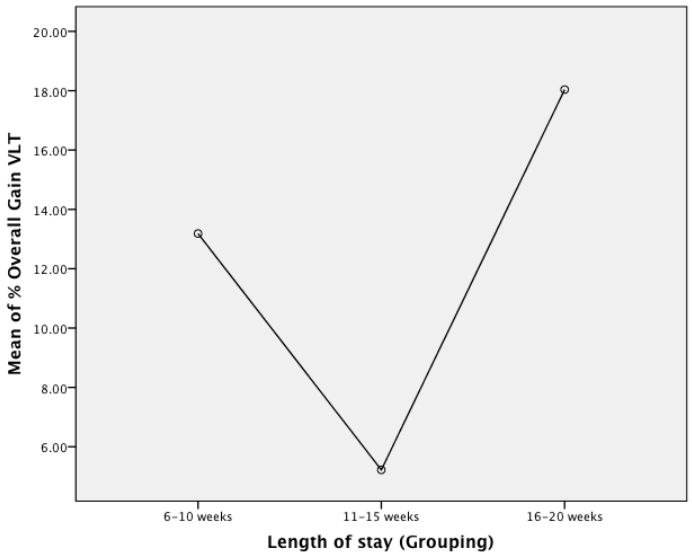
A Kruskal-Wallace test was conducted to determine whether the non-parametric findings on this non-normally distributed variable would match those of the parametric test. The Kruskal-Wallace test found a statistically significant difference between the sample as grouped by length of stay [H(2)=8.36, *p*=.015], with a mean rank of 101.46 for the short stay group, 78.27 for the medium stay group and 106.69 for the long stay group. These findings indicate that the long stay group significantly outperformed the medium stay group on receptive vocabulary gain.

A one-way between-groups ANOVA was then conducted with productive percentage vocabulary gain (logarithmically transformed) as the dependent variable and length of stay as the grouping variable. The ANOVA found no statistically significant difference in productive vocabulary gain for the three groups [F(1, 126)=1.21, *p*=.301].

Finally, the overall percentage gain variable was taken as the continuous dependent variable in a one-way between-groups analysis of variance with length of stay (short, 6-10 weeks, N=45; medium, 11-15 weeks, N=47; and long, 16-20 weeks, N=101) as the independent grouping variable. There was a statistically significant difference in overall

vocabulary gain for the three groups [F(2, 190)=3.749, *p*=.025]. Despite reaching statistical significance, the difference in mean scores between the groups was quite small. The effect size, calculated using Eta squared, was .04. Post-hoc comparisons using the Tukey HSD test indicated that the mean vocabulary gain for the medium stay group (M=5.22, SD=21.85) was significantly different (*p*=.019) from the mean gain of 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 the medium or the long stay group.

Figure 4.12: Means plot of overall gain by length of stay



As the overall vocabulary gain variable was found to be non-normally distributed, a Kruskal-Wallace test was conducted to see if the finding of this non-parametric test would match that of the ANOVA: the Kruskal-Wallace test found a statistically significant difference between the sample as grouped by length of stay [H(2)=8.80, *p*=.012], with a mean rank of 94.51 for the short stay group, 77.93 for the medium stay group and 106.99 for the long stay group. These findings indicate that the long stay group significantly outperformed the medium stay group on overall percentage vocabulary gain.

4.1.3.3 Summary of comparison of means

The sample as a whole made a statistically significant gain between time 1 and time 2 on receptive, productive and overall vocabulary knowledge as tested by the vocabulary measure in this study.

Regarding location, there was no significant difference between participants undertaking their SAE in Oxford or London on receptive, productive or overall vocabulary gain. The parametric findings were, where appropriate, supported by the findings of non-parametric analyses.

In terms of length of stay, the long stay group outperformed the medium stay group on overall and receptive vocabulary gain but the short stay group did not differ from either other group. The findings of the parametric and non-parametric tests converged in these cases. There was no statistically significant difference in productive vocabulary gain for the sample as grouped by length of stay.

4.2 INFORMAL L2 CONTACT

In this section is detailed the analysis of data and findings pertaining to the independent variable of identification with informal L2 contact as measured by part 1 of the questionnaire.

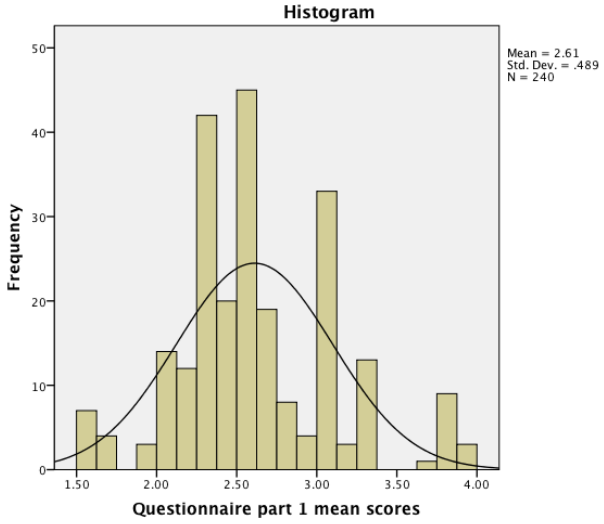
4.2.1 INTERNAL CONSISTENCY

The internal consistency of the scores in response to part 1 of the questionnaire, which comprised 28 items pertaining to informal L2 contact, was calculated using Cronbach's Alpha, finding a high coefficient of .816

4.2.2 DESCRIPTIVES AND NORMALITY

The means of the responses to the items in part 1 of the questionnaire were calculated and this variable explored with reference to normality of distribution. The mean scores on informal contact ranged from 1.50 to 3.89 (N=240, M=2.61, SD=.49). Inspection of the histogram suggested that this variable may have deviated from normality of distribution.

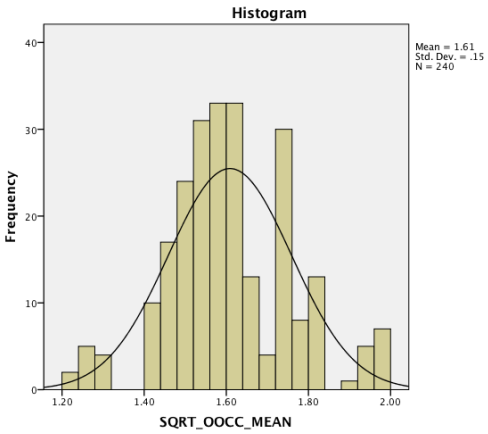
Figure 4.13: Distribution of identification with informal L2 contact



More evidence to support the observation of non-normal distribution was found in a skewness value of .509 (SE=.157) and a kurtosis value of .524 (SE=.313). The Kolmogorov-Smirnov statistic was significant at .000, further suggesting violation of the assumption of normality. No outliers were detected. Despite the non-normal distribution, the 5% trimmed mean (2.6) did not differ largely from the mean.

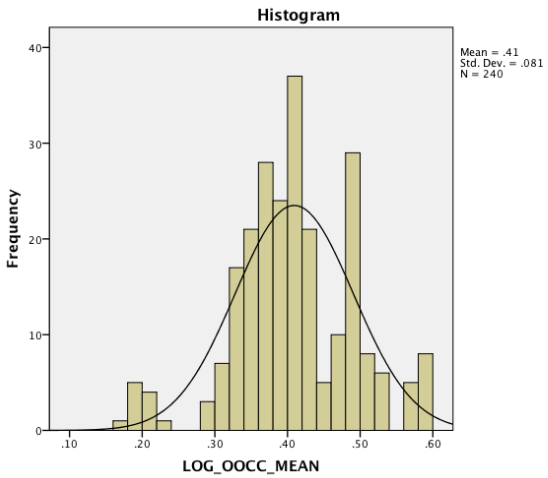
A square root transformation of the informal contact variable yielded some change in the distribution of scores: the skewness value was lessened a great deal at .183 (SE=.157) but the kurtosis value was only slightly lessened at .445 (SE=.313). The Kolmogorov-Smirnov statistic remained significant at .000.

Figure 4.14: Distribution of square root informal contact variable



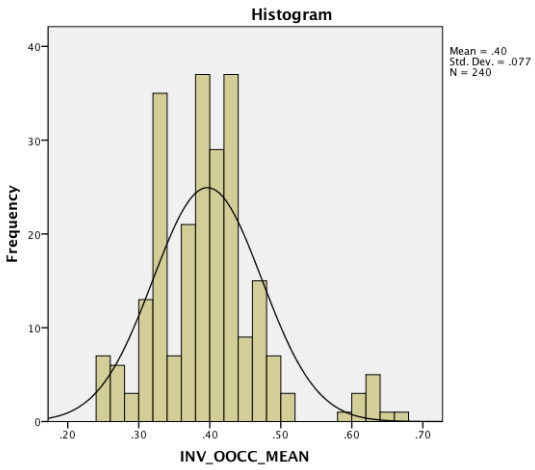
A logarithmic transformation was conducted to determine whether this approach would more normally distribute the scores, to find a similar yet negative skewness value (.119, SE=.157) and a far higher kurtosis value (-.968, SE=.312) than in the square root transformed variable, and a similarly significant Kolmogorov-Smirnov statistic.

Figure 4.15: Distribution of logarithmic informal contact variable



Finally, an inverse transformation was computed. The inversely transformed variable was even further from normality of distribution than the logarithmically transformed informal contact variables with a skewness value of .942 (SE=.157) and a kurtosis value of 1.990 (SE=.313). The Kolmogorov-Smirnov statistic remained significant.

Figure 4.16: Distribution of inverse informal contact variable



As a result of these transformations, the square root transformation of the informal L2 contact variable was retained for use in parametric analyses because it was more normally distributed than the untransformed variable and the most normally distributed of the transformations. However, as the square root transformed variable still deviated from normality, where it was used in parametric analyses a non-parametric test was also conducted.

4.2.3 DATA REDUCTION

The mean scores of the 28 questionnaire items on informal L2 contact were calculated to determine the order in which the scenarios were most identified with by the sample (shown in table 4.3) and subsequently a data reduction approach was applied to determine whether the scores in response to these items would reveal any underlying factors in the data.

Table 4.3: Order by scenario of identification with informal contact

Rank	Scenario	Mean	SD
1	Speak in English to service personnel	1.73	.800
2	Use English to get information from strangers	1.95	2.81
3	Listen to English songs	2.02	1.19
4	Read menus, timetables, posters etc.	2.12	1.07
5	Use English for short exchanges with cohabitants	2.17	1.16
6	Listen to other people's conversations	2.25	1.26
7	Listen to English movies/DVDs	2.27	.97
8	Read emails/webpages in English	2.37	1.67
9	Speak English to cohabitants	2.40	1.48
10	Speak to English teacher outside of class	2.40	1.06
11	Overall listening outside of class	2.44	1.02
12	Use English to clarify classwork outside of class	2.51	1.20
13	Listen to English TV/radio/podcasts	2.52	1.19
14	Write emails in English	2.52	1.17
15	Do long pieces of writing in English	2.54	1.00
16	Speak my L1 to native speakers of my L1	2.59	1.02
17	Read subtitles watching TV/movies in English	2.69	1.05
18	Have long conversations in English	2.70	1.22
19	Read novels in English outside of class	2.78	1.33
20	Write notes or letters in English outside of class	2.83	1.13
21	Overall writing outside of class	2.90	.99
22	Speak English with native/fluent non-cohabitants	2.90	1.26
23	Read English newspapers/magazines	2.95	1.17
24	Overall reading outside of class	2.96	1.13
25	Speak English to non-native speakers of my L1	2.99	1.09
26	Speak my L3 with native speakers of my L3	3.52	1.20
27	Speak English to native speakers of my L1	3.59	1.02
28	Speak my L1 to non-native speakers of my L1	3.70	1.20

4.2.3.1 Determining the data reduction approach

Exploratory factor analysis (EFA) was used to determine whether a similar factor structure composed of three factors (as obtained by Briggs, 2014) would be found in the present study. This application of factor analysis cannot be termed as confirmatory factor analysis (CFA) but rather acts as cross-validation of the previously found factor structure. The decision not to employ CFA was made on the basis that there existed some marked differences between the present sample and that of Briggs (2014), which may have altered the structure of the responses to the informal L2 contact items. The sample in the present study (n=241) was significantly larger than that of Briggs' (2014) study (n=95) and Briggs (2014) investigated only study abroad learners in Oxford, whereas the present study

expanded that geographical scope to include both Oxford and London. The researcher was loathe to impose on the present data a hypothesis formed in light of data which may have been representative of a slightly different population, particularly as one could argue that learners in Oxford and London may identify differently from one another with different informal scenarios, despite there being no effect of location on the vocabulary scores.

4.2.3.2 Justification for the data reduction approach

Disparity exists between researchers regarding the best approach for analysing ordinal Likert scale data (de Winter & Dodou, 2012); whether parametric (such as the *t*-test) or non-parametric tests should be used (Carifio & Perla, 2008; Jamieson, 2004). In this instance, a series of independent samples *t*-tests was conducted. However, in order to ensure that parametric analyses were indeed the correct option, a Mann-Whitney U test was conducted on the means of responses to the informal contact questionnaire items with the sample grouped by location.

4.2.3.2a *Comparison of means by location*

An independent-samples *t*-test was conducted to determine whether there was in fact a difference between the participants as grouped by location (Oxford and London) on the means of their responses to the informal contact questionnaire items. The significance level of Levene's test was p=.10 showing that the variances for the location groups were the same. There was no significant difference in identification with informal L2 contact scenarios between participants in Oxford (M=1.62, SD=.12) and participants in London (M=1.59, SD=.17; $t(238)$=1.40, p=.162). This result was mirrored by the non-significant finding of a corresponding Mann-Whitney U test [z=-1.766, p=.08].

A comparison was then carried out using the mean scores on the individual informal contact items of the sample divided into two groups based on location. The results of this second independent-samples *t*-test showed that in response to 11 of the 28 informal scenarios, the Oxford and London groups' means were significantly different from one another. A Bonferroni adjustment (0.05 / 28) was applied to find that only three of the 11 differences reached significance in this series of *t*-tests, achieved at the $p<.002$ level. The significant differences are shown in table 4.4 with a double asterisk.

Table 4.4: Location identification with informal contact

Scenario	Equal variances assumed?	F	Sig.	T	df	Sig. (2-tailed)
Read English papers/magazines	No	15.339	.000	1.947	209.383	.053
Speak to teacher	No	.553	.458	-2.579	235.048	.011
Speak to service personnel	Yes	.090	.765	-2.487	234	.014
English to clarify classwork	No	5.760	.017	-.827	212.251	.409
Read e-mails/web pages	Yes	3.845	.051	-.654	238	.514
Short exchanges with cohabitants	Yes	.200	.655	-.105	239	.916
Listen to others' conversations	No	8.395	.004	1.495	207.650	.136
Listen TV/radio/podcasts**	No	3.995	.047	5.793	226.194	.000
Speak L1 to L1 NS	Yes	1.486	.224	2.692	237	.008
Read English novels	No	5.035	.026	-2.989	235.034	.003
Converse w/native/fluent non-cohabitants	No	.476	.491	1.209	235.754	.228
Speak L1 to NNS of L1	Yes	.881	.349	-2.506	231	.013
Write emails in English	Yes	.189	.665	-2.285	239	.023
Overall reading	No	6.638	.011	.272	234.712	.786
Speak long conversations in English	No	5.820	.017	-.698	225.180	.486
Long writing in English	No	48.632	.000	-1.072	182.628	.285
Overall listening	Yes	.351	.554	2.126	239	.035
Read subtitles**	Yes	1.048	.307	3.953	239	.000
Get information from strangers	Yes	.436	.510	.019	239	.985
Read menus/posters etc.	No	7.747	.006	.238	197.266	.812
Speak Eng. to NNS of L1	No	9.405	.002	.603	189.166	.547
Listen to English movies/DVDs	Yes	.008	.928	1.497	239	.136
Write notes/letters in English	No	5.778	.017	-1.217	238.920	.225
Speak L3 with NSs of L3	No	4.362	.038	.059	217.766	.953
Overall writing**	No	22.556	.000	3.458	187.178	.001
Speak English to NSs of L1	Yes	3.476	.064	-.508	239	.612
Listen English songs	Yes	.891	.346	2.896	239	.004
Speak English with cohabitants	Yes	.569	.451	.834	239	.405

A summary of the three scenarios with which the two location groups identified differently is shown in table 4.5 The Eta squared statistic indicated a small effect size in all three cases.

Table 4.5: Summary of location identification with informal contact

Scenario	Means	SDs	Equal variances assumed?	F	Sig.	T	df	Sig. (2-tailed)	Eta^2
Listen TV/radio/ podcasts	O: 2.9070 L: 2.0636	1.08572 1.15160	No	3.99	.047	5.793	226.194	.000	0.12
Read subtitles	O: 2.9302 L:1.02459	2.4107 1.00945	Yes	1.05	.307	3.953	239	.000	0.06
Overall writing	O: 3.1163 L: 2.6696	.76677 1.16548	No	22.56	.000	3.458	187.178	.001	0.05

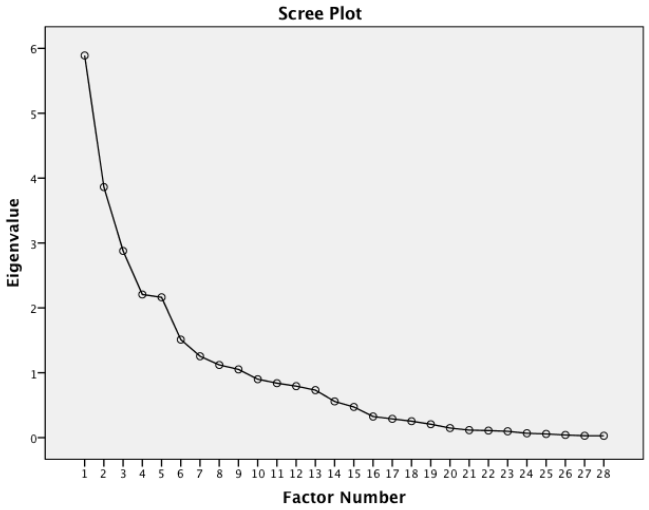
In light of these statistically significant *t*-test findings, the researcher deemed the present sample distinct from Briggs' (2014) sample and therefore decided against imposing the 2014 factor structure on the present dataset. Thus in the present study EFA, rather than CFA, was conducted on the informal contact data.

4.2.3.3 Conducting the data reduction

To ensure EFA was appropriate for the present informal contact data, the correlation matrix table was inspected and revealed numerous correlation coefficients of .3 and above. The Kaiser-Meyer-Olkin measure of sampling adequacy attained a value of .576, slightly above the recommended value of .5 (Kaiser, 1970; 1974). The Bartlett's Test of Sphericity (Bartlett, 1954) reached statistical significance (p=.000), supporting the factorability of the correlation matrix. Nine factors recorded eigenvalues above 1. These nine factors explained a total of 78.35% of the variance.

A scree plot was mapped and the factor matrix inspected in order to select which of the nine factors to retain.

Figure 4.17: Scree plot of EFA factors



The scree plot revealed a sharp elbow in the shape of the plot after factor number 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.

Table 4.6: EFA factor matrix

Factor Matrix[a]									
Scenario	Factor								
	1	2	3	4	5	6	7	8	9
Speak English with cohabitants	.791		-.314						
Overall listening	.790								
Long conversations in English	.669	-.447		.325					
Overall reading	.644							-.324	
Read menus/posters etc.	.628	.526	-.330						
Short exchanges with cohabitants	.622	-.329				.361	-.329		
Overall writing	.585				.446				
Listen to English songs	.567		.334		.520				
English to clarify classwork	.558	.312	-.320						
Get information from strangers	.317								
Speak English to NNSs of L1		.742		-.387					
Speak L1 to NNSs of L1		.623	.532						
Speak L1 to L1 NS		.610			.372				-.358
Read English novels		.595		.423					
Read English papers/magazines	.355	.575		-.338		-.339			
Speak to teacher	.469	-.515		.408	-.309				
Write emails in English			.655	.568					
Listen to English movies/DVDs			.613			-.338			-.319
Write notes/letters in English			.604	.327				.391	
Read subtitles	.406		.592		.347	-.300			
Speak English to NSs of L1	.394	.325	.334	-.418	-.324				
Listen to others' conversations	.383			-.409					
Long writing in English					.555	.535			
Speak L3 with NSs of L3	.473	.395		-.300	-.515				
Speak to service personnel				.336		-.485			.473
Listen TV/radio/podcasts	.464	-.399		-.304			.481	.325	
Read e-mails/web pages					.369		-.401		
Converse w/native/fluent non-cohabs	.496					.363	.360	-.608	

Extraction Method: Maximum Likelihood
a. 9 factors extracted.

Therefore, the first three factors were retained, and a Varimax rotation conducted better to interpret these factors.

Table 4.7: EFA varimax rotation

Total Variance Explained			
Factor	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	5.097	18.205	18.205
2	4.424	15.799	34.004
3	3.110	11.108	45.112

Extraction Method: Maximum Likelihood

4.2.3.4 Classification of the reduced data

Items that loaded onto a factor at more than .4 were analysed to determine commonalities and each factor was labelled to represent the scenarios that loaded onto it.

Table 4.8: Rotated factor matrix

Rotated Factor Matrix[a]			
Scenario	Factor		
	1	2	3
Long conversations in English	.838		
Overall listening	.791	.311	
Speak English with cohabitants	.768	.361	
Speak to teacher out of class	.673		
Short exchanges with cohabitants	.650		.353
Listen TV/radio/podcasts	.609		
Speak L1 to NNSs of L1	-.526		.534
Converse w/native/fluent non-cohabs	.431		
Overall reading	.429	.421	
Listen to others' conversations	.427		
Get information from strangers	.303		
Long writing in English			
Read menus/posters etc.		.847	
Speak English to NNSs of L1		.748	
Read English novels		.660	
Read English papers/magazines		.653	
English to clarify classwork	.304	.635	
Overall writing	.313	.608	
Speak L3 with NSs of L3		.522	.380
Speak L1 to L1 NS	-.363	.484	
Speak to service personnel		.369	
Write emails in English			.668
Read subtitles			.664
Listen to English movies/DVDs			.632
Write notes/letters in English			.625
Listen to English songs	.454		.463
Speak English to NSs of L1		.394	.461
Read e-mails/web pages			.332

Extraction Method: Maximum Likelihood
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 16 iterations.

Factor 1 comprised ten informal contact scenarios, listed below in order of weighting from heaviest to lightest and worded exactly as they appeared in the questionnaire (the scenarios that were identified with differently by the sample as grouped by location are indicated with a double-asterisk):

1. I have long conversations in English
2. Overall, I think I listen to a lot of English outside of class
3. I speak English to the people I live with
4. In England, outside of class, I speak to my teacher/instructor
5. I use English for short exchanges with the people I live with (e.g. greetings / "Please pass the salt" / "I'm leaving")
6. I listen to English TV/spoken radio/podcasts outside of class**
7. I listen to English songs outside of class
8. I speak English outside of class with native or fluent English speakers who I do not live with
9. Overall, I think I read a lot in English outside of class
10. I listen to other people's conversations in English outside of class

Nine of ten of the scenarios that loaded onto factor 1 involved speaking and/or listening to spoken language. Therefore, factor 1 was labelled 'Interactive'.

Factor 2 comprised eight informal contact scenarios weighted at .4 and above. They were, in order of heaviest to lightest:

1. I read timetables, announcements, posters, menus etc. in English outside of class
2. I speak English to non-native speakers of my native language
3. I read novels in English outside of class
4. I read English newspapers/magazines outside of class
5. I use English outside of class to clarify classroom-related work
6. Overall, I think I do a lot of writing in English outside of class**
7. I speak a language other than English or my native language to speakers of that language (e.g. Swedish with a Swedish-speaking friend)

8. I speak my native language to native or fluent speakers of my native language

Any theme running through these scenarios was less immediately obvious than with factor 1. However, one could argue that the majority of the scenarios loading onto factor 2 illustrated an application of literacy skills in English: all bar two involved reading or writing. The majority of the scenarios that did not involve reading or writing (scenarios 2, 7 and 8) neither involved the use of English. For this reason, factor 2 was labelled 'Literate'.

Factor 3 was made up of six informal contact scenarios that weighed in at .4 or higher:

1. I write emails in English outside of class
2. I read subtitles while I am watching TV or movies in English outside of class**
3. I listen to English movies and videos outside of class
4. I write personal notes or letters in English outside of class
5. I speak my native language to non-native speakers of my native language
6. I speak English to native or fluent speakers of my native language

It could be argued that the 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. As such, factor 3 was labelled 'Narrative'.

Briggs (2014) found three factors from EFA: All of the ten scenarios that loaded onto factor 1 referred to solitary contact with the L2, and therefore this factor was labelled 'Individual'. Factor 2 comprised five scenarios which all involved speaking and/or listening to spoken language. Therefore, in Briggs (2014) factor 2 was labelled

'Interactive'. All of the scenarios loading onto factor 3 related to the reception of information, either for practical purposes or for entertainment, and as a result, factor 3 was labelled 'Informative'.

The factors found in the present study in comparison with those found in Briggs (2014) are shown in table 4.9, with the heaviest weighted factor in the top row and the lightest in the bottom.

Table 4.9: Comparison with Briggs (2014) of EFA factors found

Briggs (2014)	**Present study**
Individual	Interactive
Interactive	Literate
Informative	Narrative

4.2.3.5 Exploitation of the reduced data

Three new 'factor' variables (one each for each of the factors found) were added to the dataset. The factor variables were generated by SPSS and represented the differential loadings of the informal L2 contact items onto each of the three factors respectively. These variables were computed in order to determine whether there was a relationship between them and the dependent variable, vocabulary gain. They were individually assessed for normality of distribution and found to be suitable for use in parametric analyses.

4.2.4 COMPARISON OF MEANS

4.2.4.1. ***T*-tests**

4.2.4.1a *Location and informal L2 contact*

The findings of an independent-samples *t*-test conducted to determine whether there was a difference between the participants as grouped by location (Oxford and London) on the overall sum of their responses to the informal contact questionnaire items are presented in 4.2.3.2a.

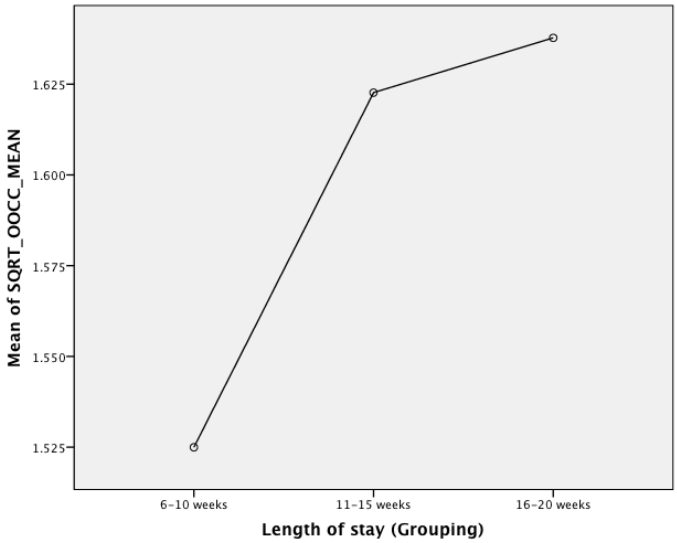
4.2.4.2 ANOVAs

4.2.4.2a *Length of stay and informal L2 contact*

To determine whether there was a statistically significant difference in identification with informal L2 contact scenarios between the sample as grouped by length of stay (short, 6-10 weeks, N=55; medium, 11-15 weeks, N=56; and long, 16-20 weeks, N=130) a one-way between-groups analysis of variance was conducted, taking the square root transformed informal contact variable (described in 4.2.2) as the dependent variable. There was a statistically significant difference in identification with informal contact scenarios for the three groups [F(2, 237)=12.06, *p*=.000]. The difference in mean scores between the groups was medium to large: the effect size, calculated using Eta squared, was .09. Post-hoc comparisons using the Tukey HSD test indicated that identification with informal L2 contact scenarios for group 1 (short stay, M=1.53, SD=.18) was significantly different from that of group 2 (medium stay, M=1.62, SD=.10) and group 3 (long stay, M=1.64, SD=.14). Groups 2 and 3 did not significantly differ from one another. According to these findings, participants in the short stay group identified more strongly (as 1=This is very true of me and 5=This is not at all true of me) with the informal L2 contact scenarios in the

questionnaire than did participants in the medium stay or long stay group. This finding was supported by a corresponding Kruskal-Wallace test, which found a statistically significant difference between the sample as grouped by length of stay [H(2)=17.13, *p*=.00], with a mean rank of 86.18 for the short stay group, 128.28 for the medium stay group and 131.41 for the long stay group.

Figure 4.18: Means plot of informal contact by length of stay



4.2.4.3 Summary of comparison of means

In terms of location, there was no significant difference in identification with informal L2 contact scenarios between participants undertaking their SAE in Oxford or in London.

Regarding length of stay, the short stay participants identified more strongly with the informal L2 contact scenarios than did the medium or long stay group. There was no significant difference between the identification with informal L2 contact scenarios of the medium stay group and the long stay group.

4.3 INFORMAL L2 CONTACT AND VOCABULARY GAIN

4.3.1 CORRELATION ANALYSES

Here follows the report of the parametric and non-parametric correlation analyses employed to determine whether there was a relationship between the independent variable of informal L2 contact and the dependent variable, vocabulary gain.

4.3.1.1 Informal L2 contact and vocabulary gain

4.3.1.1a *Informal L2 contact and receptive vocabulary gain*

A Pearson product-moment analysis was conducted on receptive vocabulary gain and the square root transformed informal L2 contact variable to find that there was no statistically significant relationship between them [r=-.023, N=196, *p*=.75]. This finding was supported by that of a supplementary non-parametric analysis [r=.034, N=196, *p*=.64].

4.3.1.1b *Informal L2 contact and productive vocabulary gain*

The relationship between the logarithmically transformed productive vocabulary gain variable and informal L2 contact variable was also explored using Pearson product-moment correlation coefficient. This analysis found no statistically significant relationship between the two [r=-.135, N=129, *p*=.20] and was supported by the finding of a corresponding Spearman analysis [r=-.166, N=129, *p*=.06].

4.3.1.1c *Informal L2 contact and overall vocabulary gain*

The relationship between overall vocabulary gain and identification with informal L2 contact was investigated using Pearson product-moment correlation coefficient. Preliminary analyses in the form of scatterplot inspection were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity, and to give an

initial impression of the relationship between the variables. There was a very slight, although not statistically significant, negative correlation between the two variables [r=-.021, N=193, *p*=.771], with higher levels of vocabulary gain associated with lower scores of identification with out-of-class contact. However, as lower scores on informal contact in fact represent higher levels of identification with the scenarios (because 1 on the Likert scale represented 'This is very true of me', and 5=This is not at all true of me) then this negative correlation was actually demonstrating a non-significant, positive (if very slight) relationship between the variables. To determine how much variance the two variables shared, the coefficient of determination was calculated, finding that identification with informal L2 contact scenarios helped to explain only .04 per cent of the variance in respondents' vocabulary gain. A Spearman correlation analysis of the same variables also failed to find a statistically significant relationship between them [r=.005, N=193, *p*=.94].

4.3.1.2 Factored informal L2 contact and vocabulary gain

4.3.1.2a *Interactive L2 contact and vocabulary gain*

The relationship between identification with L2 contact scenarios that loaded onto the Interactive factor and receptive vocabulary gain was explored using a Pearson correlation to find a very small and negative relationship [r=-.016, N=185, *p*=.826]; an insignificant finding which was supported by that of a Spearman analysis [r=.002, N=185, *p*=.974].

A Pearson correlation coefficient was then used to explore the relationship between identification with the Interactive scenarios and the logarithmically transformed productive gain variable. Again, no significant relationship was found [r=-.105, N=122, *p*=.250].

The relationship between identification with the Interactive factor and overall vocabulary gain was also investigated using Pearson product-moment correlation coefficient but no

significant relationship was found [r=.029, N=182, *p*=.699]. This finding was supported by the Spearman non-parametric alternative [r=-.003, N=182, *p*=.971].

4.3.1.2b *Literate L2 contact and vocabulary gain*

A Pearson coefficient of identification with L2 contact scenarios that loaded onto the Literate factor and receptive vocabulary gain found no significant relationship between the two [r=.052, N=196, *p*=.469] and this finding was supported by that of a Spearman analysis [r=.097, N=196, *p*=.178].

Pearson coefficient was also used to determine whether a relationship existed between identification with Literate scenarios and productive vocabulary gain. There was no significant relationship found [r=-.041, N=129, *p*=.644].

Finally, the relationship between identification with L2 contact scenarios that loaded onto the Literate factor and overall vocabulary gain was investigated using Pearson product-moment correlation coefficient. The analysis found no statistically significant relationship between the two [r=.036, N=193, *p*=.617] and this finding was bolstered by that of the Spearman non-parametric alternative [r=.064, N=193, *p*=.380].

4.3.1.2c *Narrative L2 contact and vocabulary gain*

The relationship between identification with the Narrative scenarios and receptive vocabulary gain was explored using a Pearson correlation to find no significant relationship [r=-.012, N=196, *p*=.872], a finding which was supported by a Spearman correlation [r=-.030, N=196, *p*=.677].

Similarly, no significant relationship was found between identification with the Narrative scenarios and productive gain [r=-.138, N=129, *p*=.119].

The relationship between identification with the six L2 contact scenarios that loaded onto the Narrative factor and overall vocabulary gain was also investigated using Pearson product-moment correlation coefficient, once more finding no relationship between the two [r=-.027, N=193, *p*=.710]. The findings of a Spearman correlation analysis of these variables found similar results [r=.005, N=193, *p*=.748].

4.3.1.3 Summary of findings from correlation analyses

The correlation analyses of informal L2 contact and vocabulary gain found no statistically significant relationship between the two.

4.3.2 MULTIPLE REGRESSION ANALYSES

Hierarchical multiple regression analysis was used to determine the ability of the three independent variables (informal L2 contact, location and length of stay) to predict vocabulary gain. Regression was chosen as the analytic approach so that the independent variable of primary interest (identification with informal L2 contact) could be represented on a continuous scale.

In this series of regression analyses informal contact was represented by the three factor variables as opposed to the original informal contact variable (which constituted the means of the responses to the informal contact items in the questionnaire). This decision was taken because even though no significant correlations were found between the factor variables and vocabulary gain, there may be revealed a relationship when included in the regression model with the other independent variables. The three informal contact factor variables were entered as block 1, and length of stay and location of SAE were entered together as block 2. The analyses took receptive, productive and overall vocabulary gain as the dependent variable respectively. A supplementary regression model, in which the three

factor variables were replaced in block 1 by the original informal L2 contact variable, is reported in Appendix K.

In order to investigate whether analysis of variance would support the findings of the regression analyses the informal contact variable was also dichotomised for use in a series of corresponding ANOVAs. For the ANOVAs, the informal contact variable was transformed to make a binary nominal variable using a median split based on the sum of the mean of participants' reported informal contact from the questionnaire; those with lower means rated the most number of statements as being very true of them and were thus labelled the High Identifiers (HI), and those with higher means were labelled the Low Identifiers (LI). The findings of these ANOVAs (provided in Appendix L) supported those of the regressions now described.

4.3.2.1 Sample groupings

Informal L2 contact, represented in the model by the three factored variables, was taken as the independent variable of primary interest. In terms of the other two independent variables, the sample was divided into a three-level ordinal variable based on length of stay (group 1, 6-10 weeks; group 2, 11-15 weeks; group 3, 16-20 weeks) and into a binary categorical variable based on the location of the SAE (Oxford or London).

4.3.2.2 Parametric assumptions

Here follows a description of how the variables were investigated and treated to ensure they met the parametric assumptions required for regression analysis.

4.3.2.2a *Ratio of cases to independent variables*

To decide upon the minimum number of cases necessary to conduct multiple regression analysis Tabachnick and Fidell (2001) proffer the following formula, where *m* signifies the

number of independent variables: N>50+8*m*. The receptive gain variable comprised 196 cases and the logarithmically transformed productive gain variable comprised 129 cases. The overall vocabulary gain variable comprised 193 cases. All three dependent variables therefore exceed the ratio advised by Tabachnick and Fidell (2001).

4.3.2.2b *Treatment of outliers*

A discussion of the treatment of outliers detected in the overall and productive vocabulary gain variables is given in 4.1.1.4a and 4.1.1.4b respectively. An account of the logarithmic transformation of the productive vocabulary gain variable in response to outliers is provided in 4.1.1.4c. No outlaying cases existed in any of the independent variables.

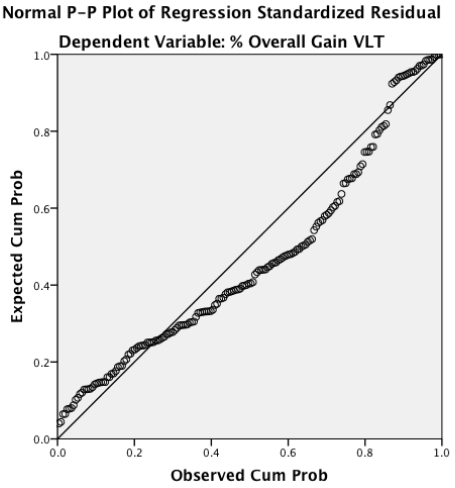
4.3.2.2c *Multicollinearity and singularity*

An initial regression was run to check the assumptions of singularity, normality, linearity and homoscedasticity. The initial regression model took overall vocabulary gain as the dependent variable. The tolerance values for the eleven independent variables as shown in the collinearity statistics output of the initial regression model were all high (.768 to .973). These values provided evidence to suggest that the assumption of the absence of multicollinearity had not been violated.

4.3.2.2d *Normality, linearity and homoscedasticity of residuals*

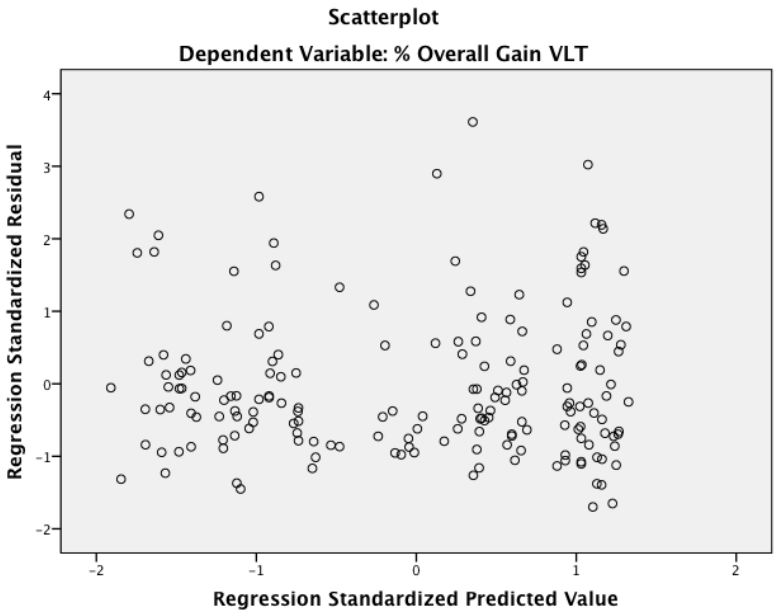
The normal probability plot generated from the initial model (figure 4.19) exhibited a reasonably straight, diagonal linear manifestation of the points, suggesting that the assumption of normality had not been violated.

Figure 4.19: Initial regression normal probability plot



The shape of the residuals scatterplot (figure 4.20) indicated that heteroscedasticity of residuals was not an issue in the model. However, a slight curve in the shape of the scatterplot suggested possible non-linearity in one or more of the variables.

Figure 4.20: Initial regression residuals scatterplot



After some investigation this curve was attributed to the non-linearity of the length of stay variable, which had initially been detected in the ANOVA run on length of stay and

overall vocabulary gain (reported in 4.1.3.2a): the long stay group (N=101, M=18.04, SD=27.73) had the highest overall gain, the medium stay group (N=47, M=5.22, SD=21.85) gained significantly less (p=.019) than the long stay group, whilst the short stay group (N=45, M=13.18, SD=28.29) gained more than the medium stay group but did not differ significantly from either other group.

As a result of the non-linearity of the length of stay variable, this variable was dummy coded into two new variables for use in the regression model, taking the long stay group as the reference because it contained the highest number of cases and was therefore the dominant category. Therefore, dummy variable 1=short stay (6-10 weeks) or any other case, and dummy variable 2=medium stay (11-15 weeks) or any other case.

4.3.2.2e *Independence of errors*

The Durbin-Watson statistic of the final model of the initial regression was 2.1, suggesting that the residuals were uncorrelated and therefore that the assumption of independence of errors had not been violated.

4.3.2.2f *Solution outliers*

In order to inspect the model for solution outliers, the Mahalanobis distances produced by the initial regression were investigated. Tabachnick and Fidell (2001) state that the critical chi-square value for a model with six independent variables (counting each dummy variable separately) is 22.46 and that any cases that exceed this value can be deemed as solution outliers. No cases in the model had a Mahalanobis value exceeding 22.46.

4.3.2.3 Receptive vocabulary gain

A hierarchical multiple regression analysis was conducted to determine the extent to which the independent variables (block 1= the three factored informal contact variables; block 2=

length of stay and location of SAE) were able to predict the receptive vocabulary gain scores. After the entry of block 1, the proportion of variance explained was 0.4%. After block 2 had been entered into the model, the model as a whole explained 3.2% of the variance. Neither model 1 [F(3, 181)=.234, *p*=.873] nor model 2 [F(6, 178)=.995, p=.430] reached statistical significance.

Length of stay was the only predictor variable to make a statistically significant contribution to the regression model (*β*=-.182, *p*=.03): this finding mirrors that of the ANOVA of these variables (reported in Appendix L), in which length of stay came close to reaching significance [F(2, 191)=2.86, *p*=.06] with a small effect size (Eta squared=.03). The negative beta value of this finding in the regression indicates that membership of the medium stay group (11-15 weeks) predicted lower receptive vocabulary gain.

Table 4.10: Predictor contribution to receptive gain regression

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	13.174	14.216		.927	.355
Interactive scenarios	-.120	.345	-.027	-.348	.728
Literate scenarios	.356	.445	.062	.801	.424
Narrative scenarios	-.146	.648	-.017	-.226	.822
2 (Constant)	13.368	15.312		.873	.384
Interactive scenarios	-.077	.373	-.017	-.206	.837
Literate scenarios	.105	.456	.018	.231	.818
Narrative scenarios	.314	.677	.037	.464	.643
Location of SAE	-.371	5.184	-.005	-.072	.943
Length of stay dummy variable 1: 6-10 wks?	-.479	6.933	-.006	-.069	.945
Length of stay dummy variable 2: 11-15 wks?	-14.892	6.823	-.182	-2.183	.030

4.3.2.4 Productive vocabulary gain

Hierarchical regression analysis was conducted to determine the extent to which the independent variables were able to predict the logarithmically transformed productive

vocabulary gain variable. After the three factored informal contact variables had been entered, the overall model explained 2.5% of the variance in productive vocabulary gain, but this finding was not statistically significant [F(3, 118)=.999, *p*=.4]. After block 2 had been entered (location and length of stay) the model as a whole explained 5% of the variance but again, this model did not reach statistical significance [F(6, 115)=1.004, *p*=.43]. None of the predictor variables made a statistically significant contribution to the final model.

4.3.2.5 Overall vocabulary gain

Hierarchical multiple regression was also conducted to determine the extent to which the six independent variables were able to predict the overall vocabulary gain scores. After block 1 had been entered the model explained 0.3% of the variance. After block 2 had been entered the model explained 3.9% of the variance. Neither model 1 [F=(3, 178)=.196, *p*=.9] nor model 2 [F(6, 175)=1.173, *p*=.32] attained significance. As in the regression analysis run with receptive gain as the dependent variable, the sole predictor variable to make a significant contribution in this model was length of stay (*β*=-2.49, *p*=.014): once more, belonging to the medium length of stay group (11-15 weeks) appeared to predict lower overall vocabulary gain. Again, the corresponding analysis of variance (Appendix L) found a supporting significant main effect for length of stay [F(2, 188)=3.62, *p*=.029] with a small effect size (Eta squared=.04). Post-hoc comparisons using the Scheffe test indicated that the mean vocabulary gain of the medium stay group (11-15 weeks; N=47, M=5.22, SD=21.85) was significantly different from the mean gain of the long stay group (16-20 weeks; N=101, M=18.04, SD=27.73).

Table 4.11: Predictor contribution to overall gain regression

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	12.224	11.143		1.097	.274
Interactive scenarios	.101	.270	.029	.371	.709
Literate scenarios	.166	.349	.037	.476	.634
Narrative scenarios	-.267	.508	-.041	-.526	.600
2 (Constant)	15.596	11.962		1.304	.194
Interactive scenarios	.052	.291	.015	.178	.859
Literate scenarios	-.051	.356	-.011	-.143	.887
Narrative scenarios	.145	.529	.022	.275	.784
Location of SAE	-1.007	4.050	-.019	-.249	.804
Length of stay dummy variable 1: 6-10 wks?	-4.310	5.416	-.067	-.796	.427
Length of stay dummy variable 2: 11-15 wks?	-13.231	5.330	-.208	-2.482	.014

4.3.2.6 Summary of multiple regression analyses

The findings of the regression analyses discussed in this chapter indicate that the model as a whole did not significantly predict vocabulary gain. Only one predictor variable in the model, length of stay, made a statistically significant contribution: belonging to the medium stay group predicted lower receptive and overall vocabulary gain but had no significant relationship to productive vocabulary gain.

4.4 SUMMARY OF INFORMAL SECOND LANGUAGE CONTACT

The top ten most identified with L2 contact scenarios were:

1. I speak in English to service personnel
2. I use English to get information from people I don't know
3. I listen to English songs outside of class
4. I read timetables, announcements, posters, menus etc. in English outside of class
5. I use English for short exchanges with the people I live with

6. I listen to other people's conversations in English outside of class
7. I listen to English movies and videos outside of class
8. I read emails/webpages in English
9. I speak English to the people I live with
10. In England, outside of class, I speak to my teacher/instructor

The short stay participants identified more strongly with the informal L2 contact scenarios than did the medium or long stay group. There was no significant difference between the location groups based on the means of their responses to the informal contact items overall. However, the location groups did respond differently to three individual informal contact items (Listen to TV/radio/podcasts, Read subtitles, and Overall writing), with Oxford-based participants identifying more strongly with these three scenarios.

Factor analysis on the informal L2 contact data revealed three underlying groupings: Interactive scenarios, which pertained to speaking and listening to the L2 in informal settings; Literate scenarios, which involved reading and writing in the L2; and Narrative scenarios, which comprised understanding or communicating stories of some kind. The factor that contained the highest number of scenarios and onto which the scenarios were most heavily weighted was the Interactive factor.

The regression analyses on factored informal L2 contact and vocabulary gain indicated that the model as a whole did not significantly predict vocabulary gain. Only one predictor variable in the model, length of stay, made a statistically significant contribution: belonging to the medium stay group predicted lower receptive and overall vocabulary gain but had no significant relationship to productive vocabulary gain.

A STUDY OF THE RELATIONSHIPS BETWEEN INFORMAL SECOND LANGUAGE CONTACT, VOCABULARY-RELATED STRATEGIC BEHAVIOUR AND VOCABULARY GAIN IN A STUDY ABROAD CONTEXT

Jessica G. Briggs

KELLOGG COLLEGE

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN EDUCATION

TRINITY TERM, 2014

VOLUME II

CHAPTER V

FINDINGS PART II

VOCABULARY-RELATED STRATEGIC BEHAVIOUR & VOCABULARY GAIN

The findings reported in this chapter pertain to the second and third research questions that the study sought to address:

2. What vocabulary-related strategies do learners of English employ in informal L2 contact in a study abroad context?
3. Is there a relationship between vocabulary-related strategy use in informal L2 contact and vocabulary gain in a study abroad context?

   3a. Do location of SAE and length of stay play a role in this relationship?

5.1 QUALITATIVE STRATEGIC BEHAVIOUR

In this section is detailed the analysis of data and findings pertaining to the independent variable of vocabulary-related strategic behaviour as gathered through the use of the Opportunities With Language Simulator (OWLS) scenarios as stimuli in semi-structured interviews. Firstly, a detailed description of the sampling strategy adopted and of the participants from whom the interview data were gathered is given. Then a description of how the interviews were transcribed is provided, followed by an account of the data analysis technique and how issues of validity and reliability pertaining to the analysis were addressed. Subsequently, the findings of the qualitative data gathered from the OWLS interviews are provided with reference to each individual simulation.

5.1.1 SAMPLING FRAME

N=36 participants comprised the qualitative subsample who were interviewed using the OWLS. As stated in 3.4, the researcher's aim was to interview 40 participants and thus to have 20 matched pairs; however, due to participant mortality, data from 18 matched pairs was gathered. In order to reduce the sample from 241 to 36 systematically, a matched pairing approach was employed. The matched pairs were selected as follows: participants were grouped according to the percentile (e.g. 1st quartile, median, 3rd quartile, 4th quartile) into which their overall vocabulary gain scores fell. Similarly for identification with informal L2 contact scenarios and vocabulary-related strategic behaviour (from the questionnaire parts 1 and 2), a mean identification score was calculated for each participant and they were grouped in the same way. A matched pair of participants scored in the same quartile for overall vocabulary gain, for identification with informal L2 contact scenarios, and for identification with vocabulary-related strategies. Moreover, a matched pair of participants was in the same group in terms of location and length of stay, and both participants in a pair had the same proficiency level. The difference between participants in a pair was their L1; one participant's native language used a Latin-based writing system and the other participant's did not.

5.1.2 PARTICIPANTS

Each matched pair was allocated an alphabetic code (i.e. pair A, pair B) and each participant in a matched pair was given a numeric code; the participant whose native tongue used a Latin writing system was allocated the number 1 (thus becoming participant A1 in pair A) and the participant whose native tongue did not use a Latin writing system became participant 2 (i.e. participant A2). N=18 participants' L1 used a Latin writing system and n=18 participants' L1 did not. L1 was employed as the point of difference

between the participants in each matched pair so that they could be matched on all of the dependent and independent variable measures in the study (vocabulary gain percentile, informal contact percentile, strategic behaviour percentile, location and length of stay). Furthermore, as post-hoc statistical analyses (reported in section 5.5) revealed that there were significant differences between the sample as grouped by proficiency level, pairs were also matched on proficiency. As such, L1 writing system was the only remaining variable to differentiate between the participants that could be dichotomised. This matched pair approach conforms to the technique posited by Tuckman and Harper (2012), who state that participants should be matched on the variables that the researcher wishes to control for.

The frequency with which different L1s were represented in this subsample is demonstrated in table 5.1.

Table 5.1: Qualitative subsample L1s

Latin writing system		Non-Latin writing system	
No. of cases	**Language**	**No. of cases**	**Language**
15	Spanish	10	Japanese
1	Czech	3	Korean
1	French	2	Mandarin
1	German	1	Arabic
		1	Russian
		1	Thai

The majority of the qualitative data subsample (9 pairs, or 50%) came from the median group, or 2nd percentile, of the overall sample in terms of their overall vocabulary gain. Four pairs (22%) came from the 3rd percentile, three pairs (17%) from the 1st percentile, and two pairs (11%) from the 4th percentile. With reference to informal L2 contact scenarios the sample were also grouped by the percentile into which the strength of their identification fell: three pairs (17%) came from the 1st percentile, five (28%) came from the median group, seven (39%) came from the 3rd percentile, and three (17%) were from

the 4th percentile. In terms of identification with vocabulary-related strategic behaviour, only one matched pair (6%) came from the 1st percentile, whereas eight pairs (44%) were from the median group, seven pairs (39%) came from the 3rd percentile, and two pairs (11%) from the 4th percentile.

With regard to location, 12 of the matched pairs were based in Oxford and six were based in London. In terms of length of stay, the majority (13 pairs, or 72%) were sampled from the long stay group (16-20 weeks), four pairs (22%) came from the medium stay group (11-15 weeks), and one pair (6%) from the short stay group (6-10 weeks). Finally, regarding proficiency level, three pairs (17%) were studying in Pre-intermediate level English language classes, eight pairs (44%) were studying in Intermediate classes, five pairs (28%) came from Upper-Intermediate classes, and two pairs (11%) from Advanced classes. None of the subsample interviewed using the OWLS instrument had a proficiency level of Mastery/C2. An overall summary of the characteristics of the subsample from whom the qualitative data were collected is provided in table 5.2.

Table 5.2: Qualitative subsample characteristics

No.	Participant code	Vocabulary gain percentile	Informal contact percentile	Strategic behaviour percentile	Location (Oxford/ London)	Length of stay (S/M/L)	Proficiency level*
1	A1	1	3	3	Oxford	L	2
2	A2						
3	B1	1	3	4	London	L	1
4	B2						
5	C1	1	2	3	Oxford	L	3
6	C2						
7	D1	2	3	3	Oxford	L	2
8	D2						
9	E1	2	4	3	London	L	2
10	E2						
11	F1	2	2	2	Oxford	L	1
12	F2						
13	G1	2	5	4	Oxford	L	2
14	G2						
15	H1	2	1	2	Oxford	M	3
16	H2						
17	I1	2	3	2	London	L	4
18	I2						
19	J1	2	1	3	Oxford	S	1
20	J2						
21	K1	2	3	2	Oxford	L	2
22	K2						
23	L1	2	3	3	Oxford	L	2
24	L2						
25	M1	3	2	2	Oxford	M	3
26	M2						
27	N1	3	4	3	London	L	4
28	N2						
29	O1	3	3	2	London	L	3
30	O2						
31	P1	3	4	2	Oxford	L	2
32	P2						
33	Q1	4	2	1	London	M	2
34	Q2						
35	R1	4	1	2	Oxford	M	3
36	R2						

* 1=Pre-intermediate/A2, 2=Intermediate/B1, 3=Upper-intermediate/B2, 4=Advanced/C1, 5=Mastery/C2

N=22 of the subsample were aged between 18-24 years old, n=10 were aged between 25-31, n=3 between 32-38, and n=1 between 39-45 years old. The subsample comprised a preponderance of female participants: the OWLS data was gathered from 21 women and 15 men.

5.1.3 TRANSCRIPTION PROTOCOL

Where a participant used a paralinguistic cue or physical gesture that served to alter or enhance the meaning of an utterance (e.g. laughing, pointing to something on the computer screen, wagging a finger), this information was italicised and included in brackets.

Contractions were transcribed as produced by the interlocutors. Standard orthography, word segmentation and spellings were used, excepting where a participant used a non-standard pronunciation of a word; in these cases, the word was transcribed as pronounced by the participant and signalled using an apostrophe before the first and after the last letter of the mispronounced word. For example, in an attempt orally to replicate to the target word 'fellow' heard in simulation 1, a participant produced the following utterance:

I'm not sure what is the meaning of fellow.

The participant pronounced 'fellow' as /ˈfelɔː/ and therefore the utterance was transcribed as follows:

I'm not sure what is the meaning of 'fellor'.

In terms of disfluent speech, partial words, repetitions and non-lexemes that functioned as fillers were all transcribed as spoken. Unintelligible utterances were marked by double parenthesis; high-inference transcriptions of near-unintelligible utterances were transcribed within double parentheses.

Each time in an interview that an OWLS simulation was played was indicated in the transcription in italics e.g. '*Simulation 1 plays*'. Similarly, where a simulation was repeated, this was also indicated in the transcription in italics.

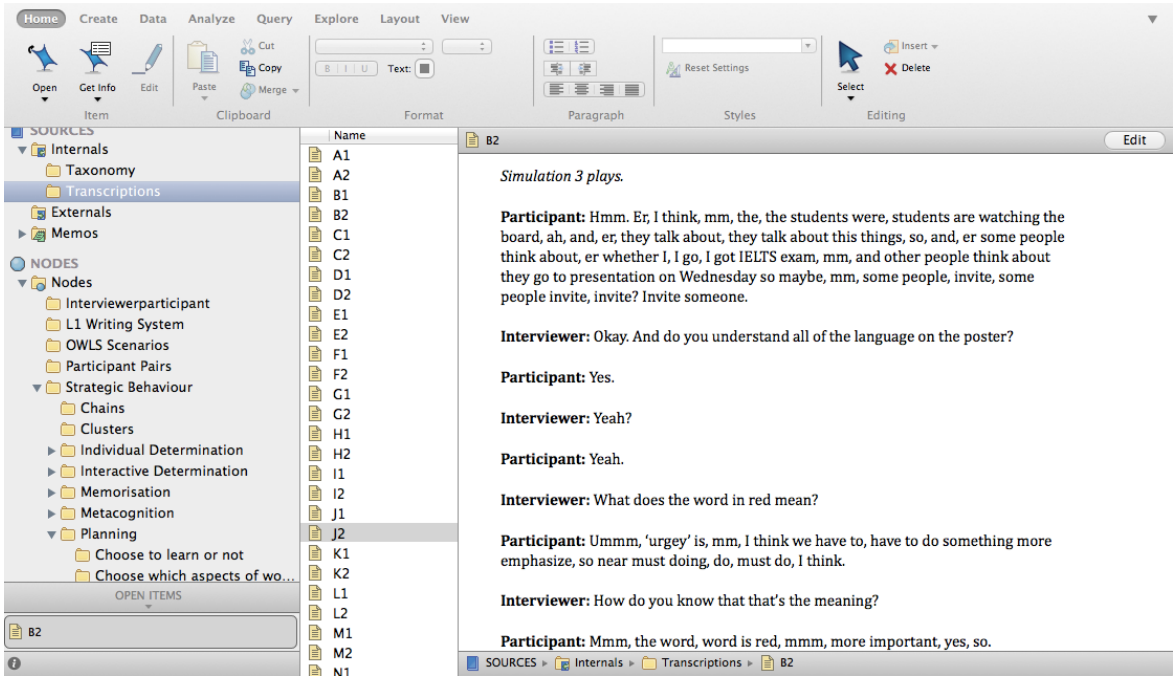
5.1.4 CODING PROCEDURE

After the transcription process was complete, the OWLS interview data was analysed. Computer-assisted qualitative data analysis software (CAQDAS) in the form of NVivo was used to aid the coding and analysis of the OWLS interview data and to create transparency by providing an electronic 'audit trail' of these processes.

The interview data was first coded into simulation nodes, whereby all of the data pertaining to each OWLS simulation was grouped together by NVivo. The data was then coded into people nodes, whereby utterances were grouped according to whether they had been said by the interviewer or the participant, by which matched pair of participants the utterance had been gathered with (e.g. pair A, pair B), and by participant A or participant B (i.e. whether a participant whose L1 used a Latin writing system had produced the utterance or a participant whose L1 did not use a Latin writing system had produced it).

The bulk of the coding that was done was with reference to strategic behaviour. The strategies taxonomy (Appendix H) was added as a source to the NVivo dataset and a parent node was created for each of the five categories of strategies in the taxonomy (planning, individual determination of meaning, interactive determination of meaning, memorisation, and metacognition). Child nodes were then created for the strategies listed in each parent node so that when an instance of strategic behaviour was coded, it would be coded not only to the specific strategy itself but also to the category into which the strategy mentioned/alluded to fell. A screen shot displaying the parent and child nodes in the NVivo dataset is shown in figure 5.1.

Figure 5.1: Screen shot of coding nodes in NVivo



On each occasion that a participant mentioned or alluded to a strategy that appeared in the taxonomy, this mention was coded as an instance of strategic behaviour. Where a strategy that was not present in the taxonomy was mentioned or alluded to, this strategy was added to the taxonomy and then similarly coded as an instance of strategic behaviour. If a participant mentioned or alluded to manifesting a combination of strategies in response to any one OWLS scenario, this was coded as strategic behaviour in one of two ways: (1) where a participant indicated that they had previously used this very same combination of strategies and therefore that they were deploying this combination of strategies systematically, an instance of strategic chaining was coded; (2) where a participant mentioned or alluded to manifesting a combination of strategies and did not indicate that they had used this same combination previously, an instance of strategic clustering was coded. Each instance of strategic behaviour, strategic chaining and strategic clustering was also coded as either high or low inference; that is to say, an instance was coded according to how clear or ambiguous the mention of/allusion to the strategy, chain or cluster was. A

percentage of high/low inference codings was calculated. The full OWLS interview transcripts from one matched pair (i.e. two participants' interview transcripts) is provided in Appendix I.

5.1.5 TRUSTWORTHINESS

In their seminal work on research into naturalistic settings, Lincoln and Guba (1985) posit that trustworthiness in qualitative research is equivalent to validity and reliability in the quantitative paradigm. They state that trustworthiness is comprised of four aspects: (1) credibility, whereby there is confidence in the truth of the reported findings of a piece of qualitative research from the perspective of the participants; (2) transferability, whereby there is confidence that the findings may be applicable to other settings; (3) dependability, whereby there is confidence that there is consistency within the findings and that very similar findings would arise if the study were to be repeated; and (4) confirmability, whereby there is confidence that the findings pertain to the 'voice' of the respondents as opposed to any possible researcher bias. The measures taken in this study to enhance the trustworthiness of the qualitative data collection and analysis are hereby discussed with specific reference to credibility, transferability, dependability and confirmability.

5.1.5.1 Credibility

In order to establish credibility, member checks were carried out with n=4 participants (10%) who came from the qualitative subsample. The four participants were emailed a copy of the transcript of their respective OWLS interview one week after the interview had taken place and were asked to read the transcript and then to reply to the researcher by email to confirm whether, in their opinion, the transcript was representative of the interview. A copy of a member check is shown in Appendix J. All four participants confirmed that the transcript was an honest, accurate and comprehensive record of their

interview. Another strategy adopted to promote the credibility of the qualitative data in this study was the use of the cyclical, iterative interview schedule (Appendix G): according to Shenton (2004), iterative questioning enhances credibility in that it promotes honest and comprehensive responses from informants.

5.1.5.2 Transferability

With regard to transferability, a purposive, maximum variation sampling approach was taken in order to try to include in the qualitative subsample representatives of all levels of the variables that had been measured quantitatively, and thus to represent in the OWLS interview data a range of experiences and competencies. That is to say, the qualitative subsample comprised the full spectrum of cases with reference to L1 writing system and to percentile scores on vocabulary gain, identification with informal L2 contact scenarios, and identification with vocabulary-related strategic behaviour. The only exception to this was proficiency level; none of the matched pairs' proficiency was Mastery/C2, yet the overall sample did include participants of this proficiency level. The thick description (Ryle, 1949; Geertz, 1973) of the OWLS interviews here provided was intended to promote the principle of transferability by giving sufficient detail to enable the reader to judge the applicability of these findings to other language learning contexts and populations.

5.1.5.3 Dependability

According to Lincoln and Guba (1985), credibility and dependability are inextricably linked to the extent that where credibility has been established, so too for the most part has dependability. However, to address specifically dependability, the processes by which the qualitative data in this study were gathered and analysed are reported in great detail and

with a level of transparency sufficient that another researcher could repeat them and likely obtain similar findings.

5.1.5.4 Confirmability

To combat the 'researcher-as-instrument' issue, 10% of the OWLS interviews (4 interviews) were transcribed twice; once by the researcher and once by an inter-rater (an experienced English language teacher and teacher trainer), with both transcribers following the protocol set out in 5.1.3. The researcher and the inter-rater both then compared these transcripts and marked any differences with a highlighter pen. This process revealed that the only differences between the transcriptions lay in the punctuation and in the spelling of non-lexeme fillers. The same inter-rater was also employed to enhance confirmability in terms of the coding process; providing agreement on the coding of mentions of/allusions to strategic behaviour, on which reports of strategic behaviour were high and low inference, and in terms of deciphering the low inference reported strategies.

5.1.5.4a *Inter-rater reliability*

Seven of the OWLS interview transcripts (20%) were coded by the researcher and by the inter-rater. Inter-rater reliability was calculated as a percentage (rather than, for example, Kappa's coefficient) because for each mention of/allusion to a strategy by a participant, there were at the outset of the coding process 46 possible strategies in the taxonomy for the raters to code to, and thus the probability of the two raters agreeing by chance was very low. Agreement between the researcher and inter-rater on the coding of strategies (that is, which strategy or chain or cluster of strategies was being referred to by the participant) stood at 79%. Agreement between the researcher and the inter-rater with regards which reports of strategic behaviour were high or low inference stood at 92%.

5.1.6 OVERALL RESPONSES TO THE OWLS

The number of instances of individual strategies by taxonomy classification, the strategic clusters and strategic chains reported in response to the eight OWLS simulations, and the total number of strategic responses to each simulation, are shown in table 5.3.

Table 5.3: Overall strategy use by taxonomy classification/combination

Sim.	Planning	Determ.: Indiv.	Determ.: Inter.	Memorisation	Metacog.	Chains	Clusters	Total
1	0	14	20	5	4	8	12	**63**
2	0	42	1	2	0	1	2	**48**
3	5	18	15	3	0	10	4	**55**
4	6	2	34	0	0	2	24	**68**
5	5	18	0	3	5	36	3	**70**
6	0	41	0	1	1	9	10	**62**
7	18	20	3	0	0	0	4	**45**
8	8	31	0	0	0	0	2	**41**
Total	**42**	**186**	**73**	**14**	**10**	**66**	**61**	**452**

Table 5.3 demonstrates that simulation 5 (listen to English songs) prompted the most strategic responses from the subsample and simulation 8 (read timetables, announcements, posters, menus etc.) the fewest. The greatest number of strategic combinations was reported with reference to simulation 5, whereas simulations 2 (listen to English DVDs/movies), 7 (use English for short exchanges with cohabitants) and 8 prompted comparatively few strategic combinations.

There now follows a description of each individual simulation and a detailed report of the strategic behaviour reported in response by the participants. For each simulation the findings with reference to individual strategy use is reported, followed by a report of the findings with reference to the use of combinations of strategies.

5.1.7 SIMULATION 1

Simulation 1 derived from questionnaire item number 6 and simulated the scenario 'Use English for short exchanges with cohabitants'. The target lexical item in simulation 1 was 'fellow'. Simulation 1 depicts a group of friends at home chatting over dinner. In the second slide of the simulation the stimulus is delivered by one of the male diners and is directed toward the participant. The man says, "So, what are your fellow students at the language school like?"

Simulation 1 prompted 63 total strategic responses, of which 43 were mentions of/allusions to individual strategies, eight to strategic chains, and 12 to strategic clusters (table 5.3). The fewest number of strategies (including strategies which comprised part of a cluster or chain) reported by any one participant in response to simulation 1 was zero, and the highest was five. The proportion of strategic responses that were coded as high inference was 14%.

5.1.7.1 Individual strategies

Of the 43 reports of individual strategy use in response to simulation 1 there were no planning strategies, 14 individual determination strategies, 20 interactive determination strategies, five memorisation strategies and four metacognition strategies. The vast majority of individual strategies reported in response to simulation 1 were with regard to determining the meaning of the target lexical item 'fellow'. The interactive determination strategies that were reported were enacted solely in the L2: they included asking for repetition of the word to aid recognition of it, immediately asking the interlocutor for the meaning of the word in English, and asking the interlocutor for a paraphrase or synonym of the word. Despite slide 1 of simulation 1 showing that there were a number of people present at the dinner (there are eight people sitting around the table), none of the

interactive determination strategies reported involved anyone other than the man who uttered the stimulus. This is interesting because the people sitting around the table are a diverse group in terms of ethnic background, so it was conceivable that some participants might assume there was a L1 speaker present.

With regard to individual determination, almost all of the strategies prompted by simulation 1 pertained to guessing the meaning from context; either from the immediate textual context (i.e. the very question in which the target item was heard) or from wider background knowledge. Most of the strategic responses coded as high inference derived from opaque reports of guessing from wider context: that is, eight participants initially responded to the stimulus by answering or rephrasing the question they (erroneously) guessed had been asked, as demonstrated in the following excerpt.

> **Participant:** *Mmmm, I think, mm, he wants to talk about, about, lives, or, yesterday things, or tomorrow friend...*
> **Interviewer:** *Yeah? Why do you think that?*
> **Participant:** *Mmm, um, mm, everyday, when I eat, where I have lunch or dinner, I always talk about, um, friends, so I think, yeah.*

An individual strategy was reported with reference to simulation 1 that had not been included in the taxonomy: the participant in the following excerpt repeated the stimulus in her head not to aid memorisation but to provide herself with another opportunity to recognise the words in the question and thus to strengthen the possibility of guessing from the immediate textual context the meaning of the target lexical item.

> **Interviewer***: And do you think you did catch the correct meaning of his question?*
> **Participant***: Yeah.*
> **Interviewer:** *You did. Okay. And what helped you to decide that?*
> **Participant:** *Erm , I repeated myself his question.*
> **Interviewer:** *In your head?*
> **Participant:** *Yeah, in my, in my head.*
> **Interviewer:** *And how did that help you? That's interesting. How did that help you?*
> **Participant:** *Umm, um, I, I can, could catch it but, and er, in student life, so, I imagine, imagined his question, yes.*

As a result of this interaction, the following strategy was added to the taxonomy under the classification of individual determination: Repeat new word/phrase in head to aid recognition.

In terms of memorisation strategies, five instances of individual strategy use were reported. They comprised two reports of repeating the target item silently 'in your head', one of repeating the item out loud, and one of imagining the spelling of the target item. With regards metacognitive strategies, one instance of setting goals for learning the new word/phrase was reported, and three instances of planning opportunities to use the word/phrase. No individual instances of planning strategies were reported.

5.1.7.2 Combinations of strategies

The clusters of strategies reported in response to simulation 1 for the most part comprised a combination of determination and memorisation strategies. The most frequently reported strategic cluster combined the use of guessing the meaning from context (an almost equal spread between guessing from wider and guessing from textual context), asking the interlocutor for the meaning of the item, and repeating the word silently 'in your head'. Other memorisation strategies that occurred in clustered responses to simulation 1 included writing down the word and its meaning later, and connecting the item to personal experience. One unique example of clustering with reference to simulation 1 involved the deployment of five strategies: a high-inference report of guessing meaning from textual context, asking for repetition of the new word/phrase to aid recognition, asking the interlocutor for a paraphrase or synonym of the item, asking for repetition of the item after the meaning had been determined, and connecting the word/phrase to personal experience. The participant who manifested this cluster explicitly stated that his deployment of interactive determination strategies was uninhibited by any desire to adhere to the norms

of social interaction; that is, he said that if he was having dinner with a group of native speakers of English, he would view the occasion primarily as an opportunity for language acquisition as opposed to an occasion for relaxation or enjoyment:

> **Participant***: I ask everything because is not everyday that I can speak with English speakers like this, in this way. I think, okay, in my country I have a dinner, I will make the conversation and eat some food, some drink, and, but if I have a chance to do these things with native English speakers then I must, I have to, to, er, use this for, to learn the language.*

The most common chain reported in response to simulation 1 involved a dyad of strategies; first guessing the meaning of the lexical item from a wider background context and then choosing (not) to work with the item. This chain prompted the addition of a strategy to the planning classification of the taxonomy: whereas previously to the present study the strategy 'choose to learn word/phrase or not' had been included in the taxonomy, the strategy 'choose to work with word/phrase' had not. The difference between these two strategies was highlighted by the chain in question: before the point at which a learner decides to learn an item, they must in naturalistic learning settings decide first whether to work with the item in order to determine its meaning, or to simply ignore it. All of the informants who manifested this chain incorrectly assumed that they had understood the stimulus sufficiently to answer the question and as such chose not to spend any time on processing any unfamiliar language.

5.1.8 SIMULATION 2

Simulation 2 derived from questionnaire item number 22 and simulated the scenario 'Listen to English DVDs/movies'. The target lexical item in simulation 2 was 'nuisance'. Simulation 2 opens on a cinema audience eating popcorn and looking up towards a screen. The 20th Century Fox theme music is playing in the background. The second slide shows the cinema screen and a scene from the movie. A boy stands in a garden; his hands are

dirty and there are dark handprints on the clean, white sheets that are hanging out to dry. In front of the boy is an adult woman facing him, with her hands on her hips. The woman gasps and then says, “What have you done? Oh, you’re being such a nuisance today!”

Simulation 2 prompted 48 total strategic responses, of which 45 were mentions of/allusions to individual strategies and one strategic cluster. Two strategic chains were reported in response to simulation 2 (table 5.3). The fewest number of strategies (including strategies comprising part of a cluster or chain) reported by any one participant in response to simulation 2 was one, and the highest was three. The proportion of strategic responses that were coded as high inference stood at 4%.

5.1.8.1 Individual strategies

Of the 45 reports of individual strategy use in response to simulation 2 there were no planning strategies, 42 individual determination strategies, one interactive determination strategy, two memorisation strategies and no metacognition strategies. Unlike simulation 1, simulation 2 provoked very uniform strategic behaviour from the participants in the OWLS interviews. The vast majority of the strategic responses to simulation 2 were instances of individual strategy use, and the vast majority or those instances were coded to the taxonomy classification of individual determination. The individual determination strategies reported were, in order of the number of times they were reported: analyse any available pictures or gestures to enhance comprehension of meaning, guess from textual context (immediate context), guess from background knowledge (wider context), and analyse part of speech to determine the function of word/phrase. The one interactive determination strategy reported was that of immediately asking an L1 speaker present for the L1 translation of the word phrase, but this strategy deployment was reported in tandem with a number of caveats:

Participant: *I am with a friend from Saudi Arabia?*
Interviewer: *Er, yeah, possibly.*
Participant: *Then maybe I ask her in Arabic what it means.*
Interviewer: *Okay. And when would you ask her?*
Participant: *Well, I would, maybe I'd ask her in that moment but maybe not. If there is a lot of action...but if I don't ask her in that moment I will forget so, I don't know.*

The two memorisation strategies that were manifested in response to simulation 2 were both reports of connecting the word/phrase to personal experience: both reports of this particular strategy involved the participant linking the word 'nuisance' to a memory of their misbehaviour as children.

5.1.8.2 Combinations of strategies

Two dyadic chains were reported in response to simulation 2. In both instances the participants reported initially analysing the visual information to aid individual determination of meaning, and then stated that they would ignore the target word because they sufficiently understood the meaning of the stimulus without hearing or understanding the meaning of 'nuisance', and as such these instances were coded to the planning strategy of choosing to work with a new word/phrase or not, which had been added to the taxonomy during the analysis of the simulation 1 data (discussed in 5.1.7.2):

Participant: *Mmm, if I couldn't hear all sentence but I, I can watch the screen so I can guess the, er, guess the situation, so I think I, I'm, I don't have to hear all sentence. So I can forget, forget this, don't worry, mmm.*

One instance of clustering was reported in response to simulation 2, and this instance comprised three distinct strategies. The strategies were: analyse any available pictures or gestures to enhance comprehension of meaning, repeat word/phrase silently 'in your head', and plan opportunities to encounter/use the word/phrase:

Participant: *Er, in the movie, you can see, you, he are very dirty, and er, I, I'm not sure what, the, the word or the word that she used but I think for the picture she's like, er, what are you, have, what, like, where have you been? Why are you like this, of dirty?*
Interviewer: *Okay. So you don't know the word she used?*

Participant: *No, I don't know because is, is...I don't know, the hands, the clothes. I think the word of the, when something is, this, I don't know this word. But I can remember this maybe.*
Interviewer: *How could you remember it?*
Participant: *Er, I don't know if I can in certain but maybe I can say 'newsir' to remember this word.*
Interviewer: *Okay, and who would you say it to?*
Participant: *To me, just, in the head because, er, it is, I am in the cinema, y'know?[Laughs]*
Interviewer: *[Laughs] Okay...*
Participant: *But maybe after the movie I can say it, I can use this word.*

Therefore, this cluster was coded as comprising a strategy each from the individual determination classification, from the memorisation classification, and from the metacognition classification.

5.1.9 SIMULATION 3

Simulation 3 derived from questionnaire item number 20 and simulated the scenario 'Read timetables, announcements, posters, menus etc.'. The target lexical item in simulation 3 was 'urge'. Simulation 3 shows a group of students looking at a noticeboard. In slide 2 of the simulation a close-up of the noticeboard is shown with a focus on one poster, which sits at the centre of the board. The poster reads, "Students who are going to take the IELTS exam at the end of this month: We urge you to come to our exam preparation classes on Wednesdays at 5pm." The word 'urge' is written in red, whereas the other text is in black.

55 total strategic responses were reported with reference to simulation 3, of which 41 were mentions of/allusions to individual strategies, 10 were coded as strategic chains, and four as strategic clusters (table 5.3). The fewest number of strategies (including strategies comprising part of a cluster or chain) reported by any one participant in response to simulation 3 was one, and the highest was five. The proportion of strategic responses that were coded as high inference was 15%.

5.1.9.1 Individual strategies

Of the 41 reports of individual strategy use in response to simulation 3 there were five planning strategies, 18 individual determination strategies, 15 interactive determination strategies, three memorisation strategies and no metacognition strategies. An almost equal number of instances of individual and interactive determination strategies were reported in response to simulation 3. Six separate individual determination strategies were reported and are hereby listed in order of the frequency with which they were reported: guessing from immediate textual context (knowledge that exam classes are important to attend); guessing from wider, background knowledge (knowledge of the high-stakes nature of the IELTS exam); analysing any available pictures or gestures to enhance comprehension of meaning (the use of red font for the target word); checking for parallels with languages other than the L1 to link with previous knowledge (referring to the similarity in orthographic and phonological form between the target word and the word 'urgent'); checking for L1 cognate/translation to transfer L1 knowledge (the similarity between the target word and the word 'urgente', which is Spanish for urgent); and using a bilingual dictionary to link to L1 knowledge (referring to electronic pocket dictionaries). A typical example of the common response to simulation 3 with reference to the similarity between 'urge' and 'urgent' is demonstrated in the excerpt below:

> **Participant:** *Ok, er, they are ask the people, asking the people, er, who wants to do the IELTS that they need to came to this class preparation. And it's very urgent to people who wants to do that.*
> **Interviewer:** *Okay. And so this...? [Points to word in red]*
> **Participant**: *Is urgent. Is urgent. Is...done quickly? Yes, quickly. Something quickly, to, to, to come quickly to this preparation class.*

With regard to interactive determination strategies, the OWLS participants reported the following three strategies, listed in order of the frequency of their mention: asking a non-native speaker present for the meaning of the word/phrase (referring to the other language

school students depicted reading the poster in slide 2 of the simulation); immediately asking a L1 speaker present for a L1 translation (referring to a fellow student present, who shared the participants' L1 and knew the meaning of the word); and asking a native speaker for the meaning of the word/phrase later (referring to asking a teacher at the next opportunity).

The instances of planning reported in response to simulation 3 derived from the manifestation of one strategy: choosing to work with the new word/phrase or not. The participants who reported this strategy stated that if they were studying for the IELTS exam, their response to the stimulus would be to go to the preparation class as advertised on the poster, and that as such, they would not spend 'unnecessary' time processing the unknown lexical item. The memorisation strategy that was manifested in response to simulation 3 was making an immediate note of the usage of the word/phrase (two reports of this strategy involved the participant taking a photograph of the poster using their camera phone, and one report involved the participant removing the poster from the noticeboard!).

5.1.9.2 Combinations of strategies

The most commonly reported chains with reference to simulation 3 comprised guessing the meaning of the target item either from textual or background knowledge, and then choosing whether or not to learn the lexical item. The participants who reported this strategy all chose not to learn the lexical item because their guess of its meaning, coupled with the information they gleaned from the textual and wider contexts, was sufficient in their eyes to fulfil the requirements in terms of attending the IELTS preparation classes. The other strategic chains reported in response to simulation 3 involved the systematic use

of dictionaries in tandem with other individual determination strategies and with memorisation strategies.

Of the four strategic clusters reported in response to simulation 3, three involved the combination of individual and interactive determination strategies. For example, in the following excerpt the participant first used the available pictorial detail (in the form of the red font) to help determine the meaning of the word, and then, based on their assumption of partial comprehension, immediately asked the (presumably) non-native speaker students present for the meaning of new word:

> **Participant:** *Mmmm, in this situation I would ask the other students, because it's obvious that is very important.*
> **Interviewer:** *How is it obvious?*
> **Participant:** *Because it's in red. And it's about, er, exams and classes and lessons, so it would be important.*

The remaining strategic cluster comprised a combination of individual determination, interactive determination, and metacognitive strategies, in that the participant reported guessing meaning from context, then asking a L1 speaker present for a L1 translation, then planning opportunities to use the new word.

5.1.10 SIMULATION 4

Simulation 4 derived from questionnaire item number 25 and simulated the scenario 'Speak in English to service personnel'. The target lexical item in simulation 4 was 'bargain'. Simulation 4 opens on a supermarket aisle, with the sounds of items being scanned in the background. Slide 2 depicts a man, who can be identified as a shop assistant because he is wearing an apron and a name badge, talking to a woman. The interlocutors each hold a product in their hands and the woman appears to be pointing to the product that she holds. She says, "And are any of these included in the half price sale?" and the man replies, "You'll need to look in our bargain section, madam".

68 total strategic responses were reported with reference to simulation 4, of which 42 were mentions of/allusions to individual strategies, two were coded as strategic chains, and 24 as strategic clusters (table 5.3). The fewest number of strategies (including strategies comprising part of a cluster or chain) reported by any one participant in response to simulation 4 was zero, and the highest was three. The proportion of strategic responses that were coded as high inference was 8%.

5.1.10.1 Individual strategies

Of the 42 reports of individual strategy use in response to simulation 4 there were six planning strategies, two individual determination strategies, 34 interactive determination strategies, no memorisation strategies and no metacognition strategies. The vast majority of individual strategies reported in response to simulation 4 were interactive determination strategies. The most frequently reported strategy in this classification was to ask the interlocutor for a repetition of the sentence in which the unknown lexical item appeared in order to aid recognition of it. A typical example of this is provided in the excerpt below.

> **Participant:** *I didn't understand, so, if I would be her, I would ask again.*
> **Interviewer:** *What would you ask? How would you phrase your question?*
> **Participant:** *'Please could you repeat your answer more slowly'.*
> **Interviewer:** *More slowly? So you would want to hear it again, even though you didn't understand the response?*
> **Participant:** *Yes because I think it's due to the fact that I didn't erm, understand the whole sentence, not the meaning of the words.*
> **Interviewer:** *I see. So you think you'd give yourself another opportunity to comprehend the whole sentence?*
> **Participant:** *Yes.*

The second most frequently reported interactive determination strategy was to ask the interlocutor for a paraphrase or synonym of the word/phrase, and the third and final interactive strategy reported was to ask the interlocutor for the meaning of the unknown word.

The individual determination strategies prompted by simulation 4 were both instances of guessing meaning from the immediate textual context. Neither of the participants who reported this strategy correctly guessed the meaning of 'bargain' and they appeared to invent what the shop assistant had said as an extension of what they had comprehended from the customer's utterance.

> **Participant:** *Mmm, mm, she, she want to buy the beans, mmm, but, mmm, she want to discount the beans and so she ask the customer to discount the beans, yes.*
> **Interviewer:** *Okay, and what did he say?*
> **Participant:** *Mmm, mm, he, he want to discount but, don't, she, he can't discount.*
> **Interviewer:** *Okay, and did you hear the words he used?*
> **Participant:** *Ah, mmm, not, I hear something but, mm, I can't catch it all.*

All of the planning strategies in response to simulation 4 pertained to choosing whether or not to work with the unknown lexical item. All of the participants who manifested this planning strategy chose not to work with the word 'bargain' and instead reported that they would either (a) thank the shop assistant for their help and continue to search the shop to see if they could find the product on sale or (b) ask the shop assistant to take them to the product in the sale.

5.1.10.2 Combinations of strategies

Simulation 4 prompted far more combinations of strategies, and particularly clustered strategic responses, than any of the other simulations. The clustered responses typically comprised one or two interactive determination strategies followed by an individual determination strategy (and in some cases then followed by a memorisation strategy). A typical example of this comprised asking the interlocutor for a repetition of the sentence in which the unknown lexical item appeared in order to aid recognition of it, followed by guessing the meaning from immediate textual context (most commonly with reference to the words 'half-price sale'), followed by asking for a repetition of the new word after the

meaning had been determined or saying the word in a sentence (i.e. where is the bargain section?).

> **Participant:** *I couldn't catch it so I will ask him, 'What did you say?'*
> **Interviewer:** *Okay, and then he repeats it for you. 'You'll need to look in our bargain section'.*
> **Participant:** *Er, I have to look for the, another, section? Did he say section?*
> **Interviewer:** *Yeah, he said bargain section.*
> **Participant:** *Bargain? Bargain. Good shopping? Ah, okay, okay, I know.*
> **Interviewer:** *Okay, so would you do anything else?*
> **Participant:** *No, I, well, I can ask him to, er, where is the bargain section, please?*

Only two strategic chains were manifested in response to simulation 4. Of these, one comprised the systematic deployment of an individual determination strategy (repeat new word/phrase in head to aid recognition) followed by an interactive determination strategy (ask interlocutor for meaning of new word/phrase), and the other comprised an interactive determination strategy (ask interlocutor for meaning of new word/phrase) followed by a planning strategy (choose to learn word/phrase or not).

5.1.11 SIMULATION 5

Simulation 5 derived from questionnaire item number 25 and simulated the scenario 'Listen to English songs'. The target lexical item in simulation 5 was 'vain'. Simulation 5 depicts a woman listening to music and wearing a pair of headphones. An instrumental section of the song 'You're So Vain' by Carly Simon is playing. Slide 2 shows a close-up of the woman's face and she appears to be concentrating on the music; her eyes are closed and she has placed her fingers onto the side of one of the headphones. The stimulus is comprised of one play of the chorus of the song. The lyrics that are heard are, "You're so vain / You probably think this song is about you / You're so vain (You're so vain) / I bet you think this song is about you, don't you, don't you?"

Simulation 5 garnered 70 total strategic responses, of which 31 were mentions of/allusions to individual strategies, 36 were coded as strategic chains, and three were coded as

strategic clusters (table 5.3). The fewest number of strategies (including strategies comprising part of a cluster or chain) reported by any one participant in response to simulation 5 was one, and the highest was six. The proportion of strategic responses that were coded as high inference was 17%.

5.1.11.1 Individual strategies

Of the 31 reports of individual strategy use in response to simulation 5 there were five planning strategies, 18 individual determination strategies, no interactive determination strategies, three memorisation strategies and five metacognition strategies. The vast majority of the individual strategic responses were coded as individual determination strategies. The most commonly reported of these was to consult a reference source other than a person or a dictionary to find the word in context; that is to say, the typical response was to search for the song by entering into a search engine the lyrics that could be comprehended in order to find and read the complete lyrics, including the unknown word contained in them. Other reported individual determination strategies included analysing the part of speech to determine the function of the unknown word and guessing from immediate textual context (vis-à-vis the line 'You probably think this song is about you'). No stand-alone instances of interactive determination were reported.

Similarly to simulation 4, the planning strategies reported in response to simulation 5 were concerned with the participant choosing whether or not to process the target lexical item.

> **Participant:** *If the sound I heard I like, I try to, to find the meaning of the word, or, but if is something is common or something that I not interesting, interested, I, I, yes, I don't care, but if I like, if I like I try to find the correct meaning and the correct words.*

One of the memorisation strategies mentioned or alluded to was with reference to connecting the target lexical item to personal experience, which is exemplified in the excerpt below from an interview with a Japanese participant:

Participant: *I don't know. Vain, like ego?*
Interviewer: *Yes!*
Participant: *Ahhh, okay, okay, I know. Negative. I think Japanese is a little vain because Japanese can, er, some Japanese make up in the train, or...*
Interviewer: *Yeah? Some British people do that!*
Participant: *Oh really? So, um, my mother said if you make up in the train at foreign countries, um, it's too rude, so I thought it is not good, not good to, but Japan is not so rude maybe, ah, I can remember this word.*

The other memorisation strategies reported were two instances of imagining the spelling of the target word, and one of saying the word aloud in isolation. Of the metacognitive strategies reported, four were instances of the participant planning opportunities to use/encounter the lexical item in future (e.g. adding the song to an iTunes playlist) and one involved setting goals for learning the item (i.e. working towards being able to sing along to the song from memory).

5.1.11.2 Combinations of strategies

Simulation 5 prompted a higher number of strategic chains than any of the other OWLS scenarios. Whilst no individual instances of interactive determination strategies were reported in response to simulation 5, interaction played an interesting role in the chains that the participants were manifesting. Many reported chains began with the participant stating that they would rewind and repeat the section of the song in which the unknown word occurred. These instances were coded as high-inference instances of interactive determination; specifically, asking for repetition of the new word/phrase to aid recognition of it. Most typically, this strategy was followed by an individual determination strategy; that of consulting a reference source to find the unknown word in context. An example of this chain is illustrated in the excerpt below.

Participant: *Uhhh, mmm, er, I will search on the internet.*
Interviewer: *Where? What website?*
Participant: *Mmm, Google. If I know the title, I, I, I search the title on the Google. Uh-huh. Or if I know the singer, mm, I search, the singer's information in the internet. Yeah.*
Interviewer: *And if you don't know the title or the singer?*

> **Participant:** *Mmm. Errr, mmm, I, I will, I will listen many times and, er, mmm, so, gradually I, I can understand the words and words, so, mm, but if I, if I can't understand the meaning of them, so I search the learn word, in the words, in the internet.*

Extensions of this dyadic chain were also reported, adding to it consultation of another reference source to find the meaning (i.e. copying and pasting the target lexical item from the lyrics online into a search engine), guessing from immediate textual context, and/or analysing any available pictures or gestures to enhance comprehension of meaning (referring to searching on YouTube for the music video which accompanies the song).

The clustered responses to simulation 5 comprised only determination strategies and all began with a report of individual determination in the form of consulting a reference source to find the target word in context (searching for the lyrics of the song online). This strategy was added to in clustered responses by asking for repetition of the new word in order to aid recognition of it (i.e. playing the song/chorus again), using a bilingual dictionary to link to L1 knowledge, and/or guessing from immediate textual context.

5.1.12 SIMULATION 6

Simulation 6 derived from questionnaire item number 8 and simulated the scenario 'Listen to English TV/radio/podcasts'. The target lexical item in simulation 6 was 'turn out'. Simulation 6 opens on a woman sitting on the sofa and watching TV whilst eating a slice of pizza. In the background the BBC News theme music plays. Slide 2 shows the TV screen, which depicts a female newsreader for 'BPP News live'. The newsreader says, "And in today's news, her majesty the queen arrived this morning in Manchester to open the new children's hospital. Many thousands of people turned out on the streets to welcome her."

62 total strategic responses were reported with reference to simulation 6, of which 43 were mentions of/allusions to individual strategies, nine were coded as strategic chains, and 10

as strategic clusters (table 5.3). The fewest number of strategies (including strategies comprising part of a cluster or chain) reported by any one participant in response to simulation 6 was one, and the highest was four. The proportion of strategic responses that were coded as high inference was 12%.

5.1.12.1 Individual strategies

Of the 43 reports of individual strategy use in response to simulation 6 there were no planning strategies, 41 individual determination strategies, no interactive determination strategies, one memorisation strategy and one metacognition strategy. Similarly to simulation 2, simulation 6 promoted a somewhat uniform set of strategic responses from the participants. The vast majority of the individual strategic responses to simulation 6 were manifestations of behaviour related to individual determination of meaning. Of these the most commonly reported strategy was that of guessing from immediate textual context: this strategy was deployed most frequently with reference to the participants hearing the verb 'turn' and building their comprehension of the stimulus based on their comprehension of either this verb as a stand alone lexical item, or their knowledge of other phrasal verbs which include this verb:

> **Participant:** *Yes, um, I think turn is maybe change the direction so, and er, I think, so turn out it means look Elizabeth queens, I think, yes.*

Another frequently manifested strategy in response to simulation 6 was that of guessing from wider background knowledge. A typical example of the deployment of this strategy is provided in the excerpt below.

> **Participant:** *Mmm. Mmm. Turn. Mmmm, I think many people mmm, made a, made a queue, er, so queen walked next to the queue.*
> **Interviewer:** *Yeah? And how did you know that?*
> **Participant:** *Eh, mmm, I, I see, I think about the situation, if I, if queen visits somewhere. And in Japan also like many people made a queue, so, I, I thought this.*

Another individual determination strategy that was prompted by simulation 6 was using any available pictures or gestures to enhance comprehension of meaning, an example of which is given in the excerpt below.

> **Participant:** *This is the news. I think that is easy to know from the news the story, the words, because, er, there is pictures, usually many pictures to explain what the person, what the person is saying. Here no, but usually.*

In terms of the memorisation strategies reported in response to simulation 6, one participant related the target lexical item to her experience of being part of a crowd which had gathered in a shopping centre in her home town in order to welcome a visiting Korean pop group. The one metacognitive strategy reported was with reference to deciding how best to learn the target word: the participant had initially confused 'turn out' with 'tell off' and in realising her error she stated that she needed to make a list of the phrasal verbs that she encountered so as not to confuse their form and meanings:

> **Participant:** *Yes! Everything is off, off, up, down [Laughs]. I have to, to make a list, to study these phrasal verbs.*

5.1.12.2 Combinations of strategies

Similarly to the individual strategies reported in response to simulation 6, the combinations manifested were comprised predominantly of individual determination strategies. The most frequently reported chain was to initially to guess from textual context and then to supplement that guess using wider background knowledge, e.g.:

> **Participant:** *Yes, it was about the Queen in Manchester and that many people came to see her at the hospital. She said turned out?*
> **Interviewer:** *Yes, she said many people turned out to see the queen.*
> **Participant:** *And it is the same, er, the same meaning as turned up? It means arriving?*
> **Interviewer:** *What makes you think that?*
> **Participant:** *It makes sense I think to have many people who arrive to see the Queen.*

The strategic clusters prompted by simulation 6 for the most part comprised a combination of individual determination and memorisation strategies. The clustered strategies reported

were predominantly cognitive, as opposed to physical, actions. The following excerpt demonstrates the only physical action reported in a cluster, the memorisation strategy of saying the target lexical item aloud in isolation. This strategy was preceded by guessing from textual context (individual determination) and followed by connecting the item to personal experience (memorisation).

> **Participant:** *Er, ah, I think it is look at, looking at the queen. She says many people are looking at the queen.*
> **Interviewer:** *Okay.*
> **Participant:** *Turn out. Mmm. Turned out. Turn out?*
> **Interviewer:** *Turned out, yep.*
> **Participant:** *I have not, mm, I have not turned, mm, turn out queen, but, uh, I was, to, er, John Mayer, do you know John Mayer?*

5.1.13 SIMULATION 7

Simulation 7 derived from questionnaire item number 6 and simulated the scenario 'Use English for short exchanges with cohabitants'. The target lexical item in simulation 7 was 'take over'. Simulation 7 shows a chopping board on which are placed various different vegetables, and a woman's hand holding a knife is slicing a mushroom. Chopping noises are heard in the background. Slide 2 pans out from the chopping board to show the woman, who is still holding the knife and slicing the mushroom, standing in a kitchen and looking directly toward the participant. The sound of a telephone ringing is heard and then the woman says, "Oh, that's my phone! Can you take over here while I answer it, please?"

Simulation 7 garnered 45 total strategic responses, of which 41 were mentions of/allusions to individual strategies and four were strategic clusters. No strategic chains were reported with reference to simulation 7 (table 5.3). The fewest number of strategies (including strategies comprising part of a cluster or chain) reported by any one participant in response to simulation 7 was one, and the highest was three. The proportion of strategic responses that were coded as high inference was 6%.

5.1.13.1 Individual strategies

Of the 41 reports of individual strategy use in response to simulation 7, there were 18 planning strategies, 20 individual determination strategies, three interactive determination strategies, no memorisation strategies and no metacognition strategies. An almost equal number of planning and individual determination strategies were reported in response to simulation 7. All of the planning strategies reported were coded as instances of choosing not to work with a word; that is to say, a large proportion of the participants understood the stimulus to mean that they should take the phone over to the woman. A typical example of this miscomprehension is provided in the excerpt below.

> **Participant:** *She wants me to take the phone to her so that she can answer it.*
> **Interviewer:** *Okay, and did you hear and understand everything she said?*
> **Participant:** *Yeah, I think so.*
> **Interviewer:** *What was the phrasal verb she used?*
> **Participant:** *Take me the phone? I don't know. Maybe...*
> **Interviewer:** *So what would you do in this situation?*
> **Participant:** *I would take to, take her the phone, and er, that's it.*

With regards the individual determination strategies reported in response to simulation 7, there were mentioned guessing from textual context, guessing from wider background knowledge, and checking for parallels with languages other than the L1 to link with previous knowledge (in this case, checking for parallels with similar sounding phrasal verbs in English). Only a very small number of cases of interactive determination strategies were reported: two of these involved asking the interlocutor for confirmation of the meaning of the stimulus, an example of which is shown below.

> **Participant:** *I say to her, ask her, do you want me to bring it to you, bring the phone to you?*

The other report of an interactive determination strategy was a case of asking the interlocutor for repetition to aid recognition.

5.1.13.2 Combinations of strategies

The common miscomprehension of the target item in simulation 7, which most commonly derived from the initial deployment of the strategy of guessing from textual context, was often boosted in a strategic cluster by using any available pictures or gestures to enhance meaning. A typical example of this cluster is provided in the excerpt below.

> **Participant:** *Take over, she is saying take over here the telephone because I am busy. My hands are full. Maybe dirty.*
> **Interviewer:** [*Laughs] It does look like her hands are full!*
> **Participant:** *Yeah, she cannot answer it for herself because she is cooking, preparing the food.*

One interesting cluster reported involved using non-verbal sounds to aid comprehension in combination with using available pictures to enhance comprehension and guessing from background knowledge.

> **Participant:** *Ummm, maybe she is my mother [Laughs] and she cooked, err, lunch or dinner, so, but the phone is 'rugging', ringing, so I have to, erm, I have to tell the mother, um, who is call me, our house, and yes, I, I gave the mmm, phone, I think.*
> **Interviewer:** *Okay, and how did you decide that this is the, the interpretation of the question?*
> **Participant:** *Oh, mmm, the sound helped us to understand this conversation, um...*
> **Interviewer:** *The sound of?*
> **Participant:** *The sound of, umm, [Makes ringing noise].*
> **Interviewer:** *Ah, okay – the ringing.*
> **Participant:** *Yes, yes, and, er, sometimes I experience this situation in Japan, so I related, I could related the sound and er, picture. Yes.*

The strategy reported in this cluster that involved using the sound of the phone to enhance comprehension was coded by both the researcher and the inter-rater as a high-inference case of an individual determination strategy; that of analysing any available pictures or gestures to enhance comprehension of meaning. However, as the strategy as listed in the taxonomy did not include sounds, the taxonomy was modified so that this strategy read as follows: Analyse any available pictures, gestures or sounds to enhance comprehension of meaning.

5.1.14 SIMULATION 8

Simulation 8 derived from questionnaire item number 20 and simulated the scenario 'Read timetables, announcements, posters, menus etc.'. The target lexical item in simulation 8 was 'pick up'. Simulation 8 opens on a man driving his car along a flat, open stretch of road. In his wing mirror is shown a car overtaking on his left. In slide 2 a roadside billboard is shown, on which there is an advertisement which reads, "Pick up a Yumm bar today. Yumm = delicious!!'

41 total strategic responses were reported with reference to simulation 8, of which 39 were mentions of/allusions to individual strategies and two were strategic clusters. No strategic chains were reported in response to simulation 8 (table 5.3). The fewest number of strategies (including strategies comprising part of a cluster or chain) reported by any one participant in response to simulation 8 was one, and the highest was three. The proportion of strategic responses that were coded as high inference was 16%.

5.1.14.1 Individual strategies

Of the 39 reports of individual strategy use in response to simulation 8 there were 8 planning strategies, 31 individual determination strategies, no interactive determination strategies, no memorisation strategies and no metacognition strategies. One of the frequently reported individual determination strategies in response to simulation 8, similarly to simulation 7, was checking for parallels with a language other than the L1 to link with previous knowledge. This strategy was manifested as confusion between the target phrasal verb and other phrasal verbs in the L2 that use either the same verb or same preposition. An example of a manifestation of this strategy that derived from a previous exposure to a similar-sounding phrasal verb is provided in the excerpt below.

Participant: *I, I, I, I saw this today.*
Interviewer: *Oh yeah? Can you remember where?*
Participant: *Yes, he said like, errr, when I was young, my teacher used to pick up me, pick up on me, no.*
Interviewer: *And what did it mean?*
Participant: *No, no it was different – sorry!*
Interviewer: *What did the teacher do?*
Participant: *No, no, it was pick on, is not pick up.*

Other commonly reported individual determination strategies in response to simulation 8 included using any available pictures or gestures to enhance comprehension of meaning (referring to the graphics on the billboard), and analysing the part of speech to determine the function of word/phrase (i.e. identifying the target lexical item as a phrasal verb). Another was guessing meaning from wider background context. The reported instances of this strategy for the most part drew on participants' knowledge of the principles of advertising, as demonstrated below.

Participant: *Yeah, pick up is the phrase, the phrasal verb. It means buy one, er, buy today a Yum bar, try it.*
Interviewer: *Okay. How do know this is the meaning?*
Participant: *It is advertising so of course is, they want you to buy the product.*
Interviewer: *Okay, and would you check to see if you were right about the meaning?*
Participant: *No, because I'm driving [Laughs] and the, because advertising is always to make you buy something.*

Indeed, even where a participant did not mention or allude to using their wider background knowledge of the purpose of advertising, this background knowledge appears to have influenced their strategic behaviour; that is to say, the instances of individual deployment of planning strategies that were reported in response to simulation 8 all pertained to choosing whether or not to process the phrasal verb, and there is evidence in the data to suggest that the unanimous choice not to process the phrasal verb in these cases was due to the presentation of the stimulus in advertisement format:

Participant: *Pick up a Yum bar today it means, mm, you should buy a chocolate today, and yes, I think, mmm.*
Interviewer: *So 'pick up' means you should buy?*
Participant: *Mmmm. Yes. It is advertisement.*

5.1.14.2 Combinations of strategies

No strategic chains and only two clusters were reported in response to simulation 8. The first cluster prompted by this simulation was a combination of guessing from background knowledge (individual determination), followed by repeating the word silently in your head to aid recognition of it (memorisation), followed by checking for parallels with languages other than the L1 (in this case, checking with the L2) to link with previous knowledge (individual determination):

> **Participant:** *Okay, this is an advertisement, to, of the chocolate Yum bar, and it says you have to buy this delicious chocolate.*
> **Interviewer:** *Okay, and are you happy with all of the language on the advertisement?*
> **Participant:** *Err, yes...I think...*
> **Interviewer:** *What are you thinking?*
> **Participant:** *Pick up, pick up. Is the same to take a person with you in your car, like, come with me in the car, no?*

The other strategic combination prompted by simulation 8 was a dyadic cluster comprised of guessing from background knowledge (individual determination) and choosing not to work with the target lexical item (planning).

5.1.15 SUMMARY OF QUALITATIVE STRATEGIC BEHAVIOUR

The qualitative data here discussed have demonstrated that a number of factors were influencing the sample's vocabulary-related strategic behaviour. Firstly, physical context was found to prohibit certain strategic actions; for example, driving along a motorway in simulation 8 prohibited dictionary-related strategic behaviour. Secondly, certain non-physical contextual features appear to have impacted on strategic action; for example, the telephone ringing in simulation 7 brought a sense of urgency which may have limited the quantity of strategic responses to the stimulus. Moreover, it appears from the data that the respondents made assumptions about the simulated scenes therein which in turn limited the types of available strategies from which they chose, e.g. in simulation 5 no interactive

strategies were reported, and thus one could argue that the sample assumed that the woman who was listening to music was alone. The weight of the data considered together serves to highlight the highly context-specific nature of strategic behaviour in response to newly encountered lexis.

5.2 QUANTITATIVE STRATEGIC BEHAVIOUR

In this section is detailed the analysis of data and findings pertaining to the independent variable of vocabulary-related strategic behaviour as measured quantitatively by part 2 of the questionnaire.

5.2.1 INTERNAL CONSISTENCY

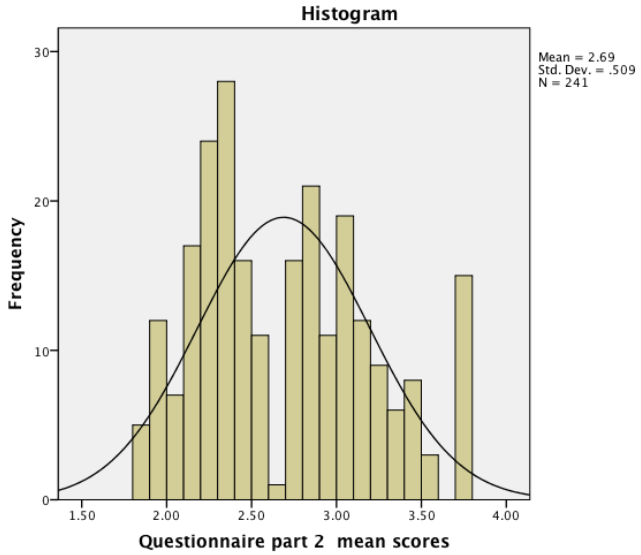
The internal consistency of the scores in response to part 2 of the questionnaire, which comprised 28 items pertaining to vocabulary-related strategic behaviour and was responded to using a 5-point 'How true of me' Likert scale, was calculated using Cronbach's Alpha, finding a high coefficient of .818.

5.2.2 DESCRIPTIVES AND NORMALITY

The means of the responses to the items in part 2 of the questionnaire were calculated and this variable explored with reference to normality of distribution. The mean scores on strategic behaviour ranged from 1.85 to 3.79 (N=241, M=2.69, SD=.51).

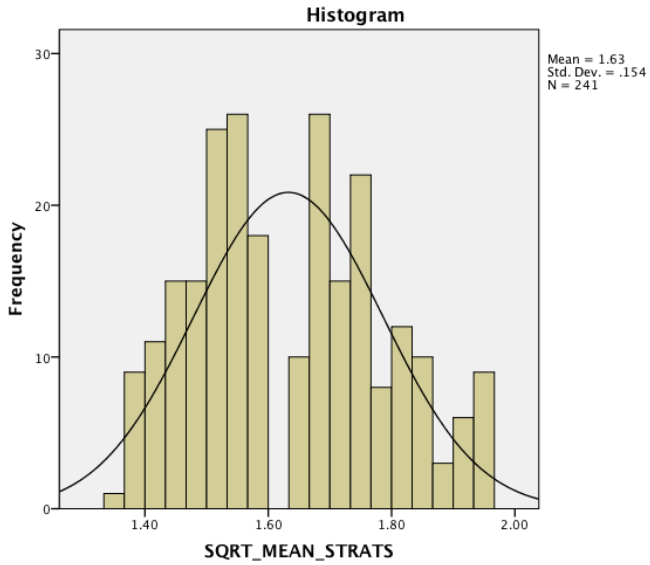
The mean scores on strategic behaviour were non-normally distributed due to a skewness value of .406 (SE=.157) and a kurtosis value of -.767 (SE=.312). The Kolmogorov-Smirnov statistic was significant at .000, further suggesting violation of the assumption of normality. No outliers were detected. Despite the non-normal distribution, the 5% trimmed mean (2.67) did not differ largely from the mean. Inspection of the histogram revealed a somewhat non-symmetrical distribution of scores with a slight positive skew.

Figure 5.2: Distribution of strategic behaviour variable



A square root transformation of the strategic behaviour variable yielded little change in the distribution of scores: the skewness value was lessened slightly at .262 (SE=.157) but the kurtosis value increased to -.900 (SE=.312). The Kolmogorov-Smirnov statistic remained significant at .000.

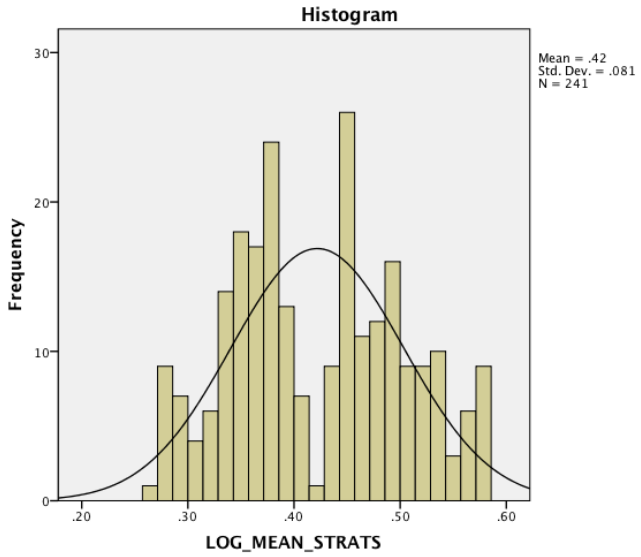
Figure 5.3: Distribution of square root transformation of strategic behaviour variable



A logarithmic transformation was conducted to determine whether this approach would more normally distribute the scores, to find very similar skewness (.119, SE=.157) and

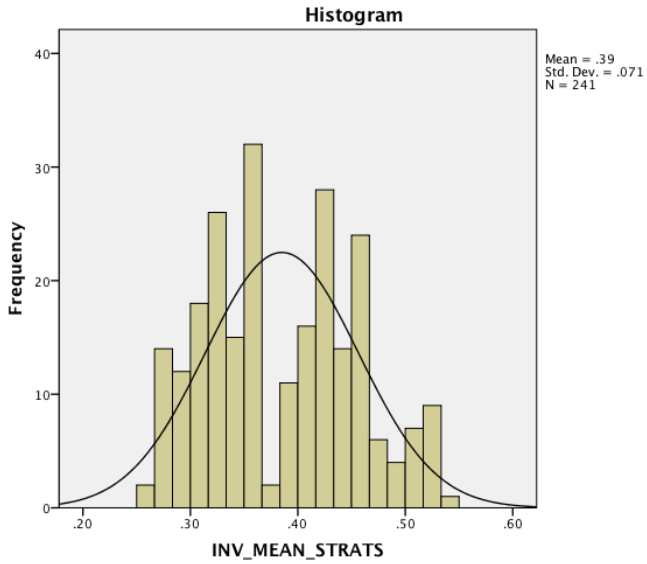
kurtosis values (-.968, SE=.312) as in the original and square root transformed variables, and a similarly significant Kolmogorov-Smirnov statistic.

Figure 5.4: Distribution of logarithmic transformation of strategic behaviour variable



Finally, an inverse transformation was made. The inversely transformed variable was the closest version of the variable to normality of distribution, yet was still non-normally distributed with a skewness value of .166 (SE=.157) and a kurtosis value of -.937 (SE=.312). The Kolmogorov-Smirnov statistic remained significant.

Figure 5.5: Distribution of inverse transformation of strategic behaviour variable



In light of the application of the data transformation investigations here described, the original strategic behaviour variable was retained for use in parametric analyses because none of the transformations succeeded in avoiding violation of the assumption of normality of distribution.

5.2.2.1 Summary of descriptives and normality

The explorations into the normality of the distribution of the strategic behaviour variable demonstrated that normality of distribution was not possible to attain. In order to account for the non-normal distribution of the original strategic behaviour variable, on each occasion that a parametric statistical analysis was carried out using this variable the equivalent non-parametric test was also conducted and reported in order to demonstrate the convergence or divergence of the findings of the two approaches.

5.2.3 DATA REDUCTION

The mean scores of the 28 items from part 2 of the questionnaire (strategic behaviour) were calculated to determine the order in which the strategies were most identified with by the sample (shown in table 5.4) and subsequently a data reduction approach was applied to determine whether the scores in response to these items would reveal any underlying factors in the data.

Table 5.4: Order by strategy of identification with strategic behaviour

Rank	Strategy	Mean	SD
1	Use a bilingual dictionary to find the L1 translation of the word/phrase	1.73	1.15
2	Repeat the word/phrase in my head to try to remember it	2.10	.99
3	Try to think of a translation of the word/phrase in my native language	2.15	1.12
4	Ask the speaker to explain (in English) the meaning of the new word/phrase	2.19	1.05
5	Try to guess the meaning from looking at the word	2.24	.83
6	Try to remember the word/phrase to ask teacher/friends later for the meaning	2.28	1.35
7	Write down the word/phrase and its meaning immediately	2.36	1.21
8	Try to decide the best way to learn the word/phrase	2.43	1.12
9	Try to connect the new word/phrase to a personal experience I have had	2.49	.93
10	Try to think of situations where I could use the new word/phrase	2.56	1.76
11	Use an English-only dictionary to find the meaning of the word/phrase	2.56	1.04
12	Try to imagine how the word/phrase is spelt	2.57	2.97
13	Try to think of a picture in my head to help me remember the meaning	2.63	1.30
14	Write down how the word/phrase was used (the whole sentence) immediately	2.65	1.25
15	Try to use the new word/phrase in a sentence as soon as possible	2.68	1.04
16	Ask the speaker to give me a synonym of the word/phrase in English	2.72	1.16
17	Try to think of English words/phrases that have a similar/opposite meaning	2.73	.93
18	Try to work out what part of speech the word/phrase is (e.g. noun/verb)	2.75	1.16
19	Think about what words could be used before and after the word/phrase	2.83	1.17
20	Use available pictures/body language to help me understand the word/phrase	2.84	1.04
21	Use an English-only dictionary to find the word/phrase used in a sentence	2.89	1.06
22	Decide whether or not I want to remember/learn the word/phrase	2.98	1.19
23	Write down the word/phrase and its meaning later	3.07	1.21
24	Say the word/phrase aloud to help me remember it	3.08	1.22
25	Ask the speaker to translate the word/phrase into my native language	3.10	1.25
26	Write down how the word/phrase was used (e.g. the whole sentence) *later*	3.30	1.19
27	Set myself goals for learning/using the word/phrase	3.43	.95
28	Draw a picture/diagram to help remember the word/phrase and its meaning	4.05	1.08

5.2.3.1 The data reduction approach

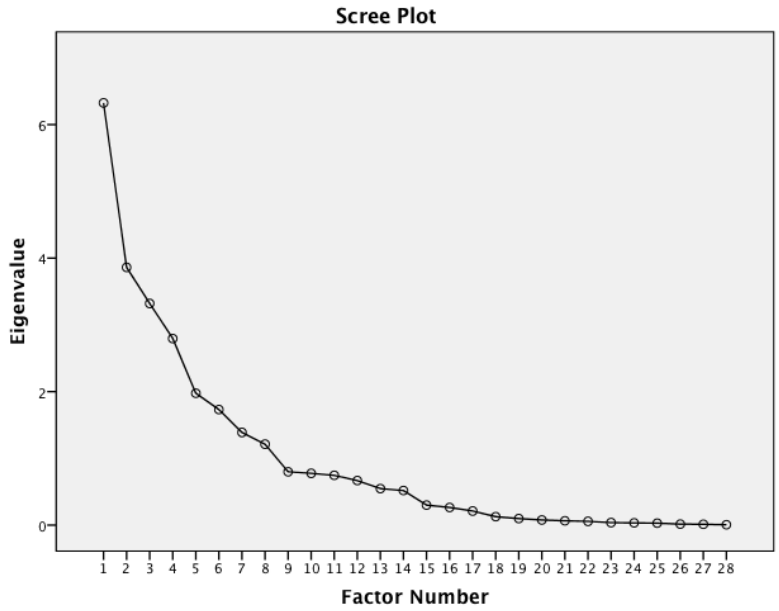
In order to match the data reduction approach taken with reference to the informal L2 contact questionnaire data (described in 4.2.3.1), the strategic behaviour questionnaire data was subjected to exploratory factor analysis.

To ensure EFA was appropriate for the strategic behaviour data, the correlation matrix table was inspected and revealed numerous correlation coefficients of .3 and above. The Kaiser-Meyer-Oklin measure of sampling adequacy attained a value of .538, slightly above the recommended value of .5 (Kaiser, 1970; 1974). The Bartlett's Test of Sphericity (Bartlett, 1954) reached statistical significance (p=.000), supporting the factorability of the

correlation matrix. Eight factors recorded eigenvalues above 1, and these eight factors explained a total of 80.75% of the variance.

A scree plot was mapped and the factor matrix inspected in order to select which of the eight factors to retain.

Figure 5.6: Scree plot of EFA factors



There appeared two elbows in the shape of the plot: one after factor number 9 and the other after factor number 5. Reference to the factor matrix revealed that only the first five factors comprised a majority of items that loaded onto them at more than .4.

Table 5.5: EFA factor matrix

Factor Matrix[a]								
Strategy	Factor							
	1	2	3	4	5	6	7	8
Work out part of speech of item	.805					-.325		
Ask for explanation in English	.763							
Think of situations to use item	.675		.346					
Remember item and ask for meaning later	.659		.315		.422			-.337
Guess the meaning of the item	.625		-.372	.394				
Write down how is item used immediately	.606					.480		
Ask for L1 translation	.596			.449				
Use L2-only dictionary for example in context	.592	.481			-.402			
Ask for synonym in English	.588		-.444	-.401				
Repeat item in head to remember	.572			-.468				
Connect item to personal experience	.544	-.536					-.429	
Use bilingual dictionary for L1 translation	.543					.521	-.412	
Write down item and meaning later		.643			.399			.394
Use item in new sentence asap	.433	-.588	.309					
Use L2-only dictionary to find meaning	.462	.571			-.499			
Use visual cues to determine meaning		.567	.515					-.399
Try to think of L1 translation		.557	.383				-.512	
Think of L2 synonyms/antonyms	.481	-.520						
Set goals to learn/use item		-.501			.493		.331	
Think of words that go before/after item		-.476			-.402	.303		
Decide (not) to learn item			-.777					
Write down how is item used later			.667			.324		.403
Decide how to learn item	.388		-.583	.324				
Imagine a picture to remember the meaning	.397		.577		.338			
Draw picture to remember item and meaning		.474		.763				
Write down item and meaning immediately		-.315		.597		.357		
Say item aloud to aid memory of it	.392			.470		-.459		.326
Imagine how item is spelled				-.445				

Extraction method: Maximum Likelihood
a. 8 factors extracted

Therefore, the first five factors were retained, and a Varimax rotation conducted better to interpret these factors.

Table 5.6: EFA varimax rotation

Total Variance Explained			
	Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %
1	4.462	15.935	15.935
2	4.268	15.244	31.180
3	3.511	12.541	43.720
4	3.039	10.853	54.574
5	2.997	10.703	65.276

Extraction Method: Maximum Likelihood

5.2.3.2 Classification of the reduced data

Items that loaded onto a factor at more than .4 in the rotated factor matrix were analysed to determine commonalities and each factor was labelled to represent the strategies that loaded onto it.

Table 5.7: Rotated factor matrix

Rotated Factor Matrix[a]					
	Factor				
	1	2	3	4	5
Use item in new sentence asap	.818				
Connect item to personal experience	.767				
Think of situations to use item	.761				
Think of L2 synonyms/antonyms	.704				
Repeat item in head to remember	.647		.305		
Remember item and ask for meaning later	.522			.486	.342
Imagine a picture to remember the meaning	.516			.441	.392
Imagine how item is spelled	.450				
Guess the meaning of the item		.833			
Ask for synonym in English		.818			
Decide (not) to learn item		.782			
Work out part of speech of item		.769			
Ask for explanation in English		.657	.531		
Decide how to learn item		.652		.381	
Use L2-only dictionary for example in context			.862		
Use L2-only dictionary to find meaning			.860		
Use visual cues to determine meaning			.708		.389
Write down item and meaning immediately				.716	
Draw picture to remember item and meaning	-.379		.483	.693	
Set goals to learn/use item	.317		-.443	.568	
Ask for L1 translation				.483	-.472
Say item aloud to aid memory of it			.453	.478	
Use bilingual dictionary for L1 translation		.407		.468	
Write down item and meaning later		.301			.706
Think of words that go before/after item					-.706
Think of L1 translation			.303		.685
Write down how is item used later					.640
Write down how is item used immediately	.361	.315			-.415

Extraction method: Maximum Likelihood. Rotation method: Varimax with Kaiser Normalisation [a]
a. Rotation converged in 14 iterations.

The first factor comprised the following eight strategies, listed in order of heaviest to lightest and worded exactly as they appeared in the questionnaire:

1. Use the item in a sentence as soon as possible
2. Connect the item to a personal experience I have had
3. Think of situations in which to use the item
4. Think of words/phrases in English that have a similar/opposite meaning
5. Repeat the item in my head to help me remember it
6. Remember the word/phrase to ask someone later for the meaning
7. Think of a picture in my head to help remember the meaning
8. Imagine how the item is spelled

All eight of these strategies presuppose that the learner has decided to try to acquire and/or use the lexical item in question. The heaviest-weighted strategy involves use of the item at the earliest possible opportunity, whilst the following seven comprise preparation for using the item. Interestingly, all eight strategies, at the point of their inception and deployment, are solitary and cognitive; that is, these are all mental actions as opposed to physical manifestations of strategic intention. As a result, this factor was labelled 'Cognitive Action in Time', henceforth referred to as 'Cognitive Action'.

The second factor was made up of six strategies:

1. Try to guess the meaning
2. Ask the speaker to give me a synonym of the item in English
3. Decide whether or not I want to remember/learn the item
4. Work out what part of speech the item is (e.g. noun/verb/adjective)
5. Ask the speaker to explain (in English) the meaning of the new item
6. Decide the best way to learn the word/phrase

In this set of strategies, it appears that the learner is making judgments about the lexical item in terms of what it means and whether/how they should learn it. As such, this factor was labelled 'Appraisal'.

The third factor comprised three strategies:

1. Use an English-only dictionary to find the item used in a sentence
2. Use an English-only dictionary to find the meaning of the item
3. Use any available pictures/body language to help me understand the item

All the strategies in this factor describe a solitary endeavour that makes use of visual information about the item in order better to determine its meaning and use. Therefore this factor was labelled 'Visual Reference'.

The fourth factor was made up of six strategies:

1. Write down the item and its meaning immediately
2. Draw a picture/diagram to help me the remember the item and its meaning
3. Set myself goals for learning/using the item
4. Ask the speaker to translate the item into my native language
5. Say the item aloud to help me remember it
6. Use a bilingual dictionary to find the translation of the item in my language

Here the learner seems to be establishing a basis from which to put the lexical item to work; acquiring the necessary knowledge base from which to work with the item further. Moreover, one could argue that these strategies would all (with the exception perhaps of strategy 3) be deployed soon after first exposure to the item. As a result, this factor was labelled 'Establishment'.

Finally, the fifth factor was comprised of five strategies:

1. Write down the item and its meaning later
2. Think about what words could be used before and after the item in sentences
3. Think of a translation of the item in my native language
4. Write down how the item was used (e.g. the whole sentence) later
5. Write down how the item was used (e.g. the whole sentence) immediately

In this factor the learner appeared to be figuring out the linguistic contexts in which the lexical item could be used and how they could fit the item into their existing schemata. As such, this factor was named 'Context Provision'.

In summary, EFA found five factors underlying the questionnaire data on identification with strategic behaviour. The five factor labels are shown in table 5.8, with the heaviest weighted factor in the top row of the table and the lightest at the bottom.

Table 5.8: Strategic behaviour factor labels

Factor
Cognitive Action
Appraisal
Visual Reference
Establishment
Context Provision

5.2.3.3 Exploitation of the reduced data

Five new 'factor' variables (one each for each of the factors found) were added to the dataset. The factor variables represented the differential loadings of the responses to the vocabulary-related strategic behaviour items onto each of the five factors respectively. These variables were computed in order to determine whether there was a relationship between them and the dependent variable, vocabulary gain.

The five factor variables were individually assessed for normality of distribution. Only one of the five factored strategic behaviour variables, Cognitive Action, contained an outlaying case, and therefore this outlier was removed from the dataset. The factor variables were normally or near-normally distributed and as such were deemed suitable for use in parametric analyses.

5.2.4 COMPARISON OF MEANS

5.2.4.1. ***T*-tests**

5.2.4.1a *Location and strategic behaviour*

An independent-samples *t*-test was conducted to determine whether there was a difference between the participants as grouped by location (Oxford and London) on the means of their responses to the strategic behaviour questionnaire items. The significance level of Levene's test was p=.071, which demonstrated that the assumption of equal variances had not been violated. There was a significant difference in identification with vocabulary-related strategic behaviour between participants in Oxford (N=129, M=2.62, SD=.48) and participants in London (N=112, M=2.77, SD=.53; $t(239)$=-2.32, p=.02). This result was mirrored by the significant finding of a corresponding Mann-Whitney U test [z=-2.68, p=.007). As the mean value of the Oxford group for this variable was lower than that of their counterparts in London, this finding indicates that participants in Oxford identified more highly with the vocabulary-related strategic behaviour (because 1=This is very true of me and 5=This is not at all true of me). The magnitude of the differences discerned by the *t*-test was small (η^2=.022).

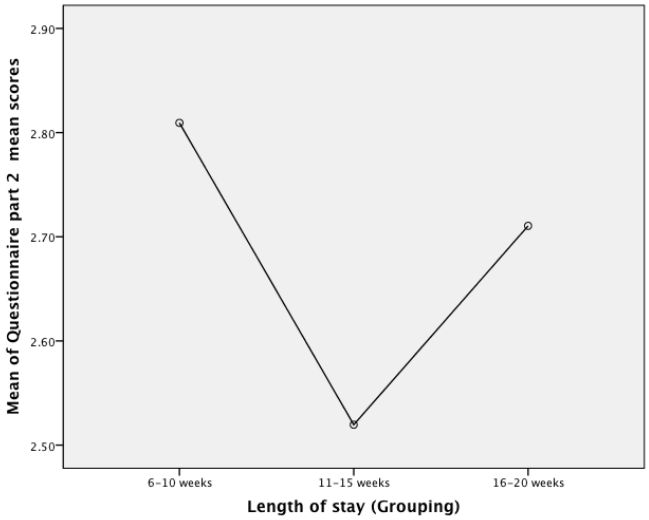
5.2.4.2 ANOVAs

5.2.4.2a *Length of stay and strategic behaviour*

To determine whether there was a statistically significant difference in the identification with vocabulary-related strategic behaviour between the sample as grouped by length of stay (short, 6-10 weeks, N=55; medium, 11-15 weeks, N=56; and long, 16-20 weeks, N=130) a one-way between-groups analysis of variance was conducted, taking the mean responses to the twenty-eight strategic behaviour items from part 2 of the questionnaire as the dependent variable. There was a statistically significant difference in identification with vocabulary-related strategic behaviour for the three groups [F(2, 238)=4.92, *p*=.008). Despite reaching statistical significance, the difference in mean scores between the duration groups was only small to medium: The effect size, calculated using Eta squared, was .04. A corresponding Kruskal-Wallis test also found a significant difference between the sample identification with strategic behaviour as grouped by length of stay (Chi-square(2)=9.33, *p*=.009).

Post-hoc ANOVA comparisons using the Tukey HSD test indicated that the mean scores on identification with strategic behaviour for the short stay group (M=2.81, SD=.40) were significantly different from those of the medium stay group (M=2.52, SD=.44; *p*=.007). The long stay group (M=2.71, SD=.56) did not significantly differ from the short stay group but differed significantly from the medium stay group (*p*=.047). According to these findings, participants in the medium stay group identified more strongly with the strategic behaviour items in the questionnaire than did participants in the short stay and long stay groups (because 1=This is very true of me and 5=This is not at all true of me).

Figure 5.7: Means plot of identification with strategic behaviour by length of stay



5.2.4.3 Summary of comparison of means

In terms of location, there was a significant difference in identification with vocabulary-related strategic behaviour between participants undertaking their SAE in Oxford and in London, with participants in Oxford identifying more highly with the items in part 2 of the questionnaire than their counterparts in the capital.

Regarding length of stay, the medium stay participants identified more strongly with the strategic behaviour items than the other two groups. There was no significant difference between the identification with strategic behaviour of the long stay group and the short stay group.

5.3 STRATEGIC BEHAVIOUR AND VOCABULARY GAIN

5.3.1 CORRELATION ANALYSES

Here follows the report of the parametric and non-parametric correlation analyses employed to determine whether there was a relationship between the independent variable of strategic behaviour and the dependent variable, vocabulary gain.

5.3.1.1 Strategic behaviour and vocabulary gain

5.3.1.1a *Strategic behaviour and receptive vocabulary gain*

A Pearson product-moment analysis was conducted on the receptive vocabulary gain and strategic behaviour variables, to find a small, positive, statistically significant relationship between the two [r=.243, N=196, *p*=.001]. A supplementary Spearman analysis also identified a significant relationship between the variables [r=.277, N=196, *p*=.000]. Lower scores on strategic behaviour represented higher levels of identification with the strategies because 1=This is very true of me and 5=This is not at all true of me, and therefore the positive correlation actually demonstrated a significant but small, negative relationship between the variables, with higher levels of identification with strategic behaviour associated with lower receptive vocabulary gain. The coefficient of determination indicated that identification with strategic behaviour explained 7.67% of the variance in the sample's receptive vocabulary gain.

5.3.1.1b *Strategic behaviour and productive vocabulary gain*

The relationship between the logarithmically transformed productive vocabulary gain variable and strategic behaviour was also explored using Pearson product-moment correlation coefficient. There was no statistically significant relationship between the two [r=-.044, N=129, *p*=.620].

5.3.1.1c *Strategic behaviour and overall vocabulary gain*

A Pearson product-moment analysis was conducted on the overall vocabulary gain and strategic behaviour variables to find a small yet statistically significant, positive relationship between them [r=.231, N=193, *p*=.001]. This finding was supported by that of a supplementary non-parametric analysis [r=.262, N=193, *p*=.000]. As lower scores on

strategic behaviour in fact represented higher levels of identification with the strategies, the positive correlation was demonstrating a significant but small, negative relationship between the variables, with higher levels of identification with strategic behaviour associated with lower overall vocabulary gain. The coefficient of determination indicated that identification with strategic behaviour helped to explain 5.3% of the variance in participants' overall vocabulary gain.

5.3.1.2 Factored strategic behaviour and vocabulary gain

5.3.1.2a *Cognitive Action strategies and vocabulary gain*

The relationship between identification with strategies that loaded onto the Cognitive Action factor and receptive vocabulary gain was explored using a Pearson correlation to find a small, positive, statistically significant correlation [r=.169, N=196, *p*=.018]. A corresponding Spearman analysis supported this finding [r=.181, N=195, *p*=.011]. The coefficient of determination indicated that identification with strategies that loaded onto the Cognitive Action factor explained 2.9% of the variance in the receptive vocabulary gain scores.

The Pearson statistic was also used to investigate the relationship between the Cognitive Action strategies and the logarithmically transformed productive vocabulary gain variable. The analysis found no significant relationship between the two [r=.001, N=129, *p*=.994].

Finally, an investigation of the relationship between the Cognitive Action strategies and overall vocabulary gain was carried out using the Pearson statistic, to find a small yet highly statistically significant, positive relationship [r=186, N=192, *p*=.010]. This finding was supported by a corresponding Spearman analysis [r=.172, N=192, *p*=.017]. The coefficient of determination indicated that identification with the strategies included in the

Cognitive Action factor accounted for 3.5% of the variance in the overall vocabulary gain scores.

5.3.1.2b *Appraisal strategies and vocabulary gain*

A Pearson product-moment analysis was conducted on the Appraisal factor strategic behaviour and the receptive vocabulary gain variables respectively. The analysis revealed no significant relationship between the two [r=.095, N=196, *p*=.187], a finding which was supported by a corresponding Spearman analysis [r=.135, N=196, *p*=.06].

In terms of the relationship between the Appraisal strategies and productive vocabulary gain, the findings were once more non-significant [r=.017, N=129, *p*=.847]. Neither was there any significant relationship found between the Appraisal strategies and overall vocabulary gain from either Pearson [r=.058, N=193, *p*=.422] or Spearman [r=.126, N=193, *p*=.08] analysis.

5.3.1.2c *Visual Reference strategies and vocabulary gain*

The relationship between identification with strategies that loaded onto the Visual Reference factor and receptive vocabulary gain was explored using a Pearson correlation to find no statistically significant correlation [r=-.048, N=196, *p*=.506]. A supplementary Spearman analysis supported this finding [r=-.073, N=196, *p*=.308].

Regarding the Visual Reference strategies and productive vocabulary gain, there was no significant correlation [r=.013, N=129, *p*=.881]. Neither was there a significant relationship between the Visual Reference strategies and overall receptive gain from either a Pearson [r=-.059, N=193, *p*=.413] or a Spearman [r=-.034, N=193, *p*=.641] analysis.

5.3.1.2d *Establishment strategies and vocabulary gain*

A Pearson product-moment analysis was conducted on the Establishment factor strategic behaviour and the receptive vocabulary gain variables. There was no significant relationship between the variables from either a parametric [r=.063, N=196, *p*=.381] or a non-parametric [r=-.008, N=196, *p*=.911] analysis.

In terms of the Establishment strategies and the logarithmically transformed productive gain variable, the Pearson analysis found a small yet statistically significant, negative relationship [r=-.190, N=129, *p*=.031]. The coefficient of determination indicated that identification with the strategies that loaded onto the Establishment factor served to explain 3.6% of the variance in the productive vocabulary gain scores.

Finally, an investigation of the relationship between the Establishment strategies and overall vocabulary gain was carried out using the Pearson statistic, to find no significant relationship between the two [r=-.038, N=193, *p*=.597]. Neither did the corresponding Spearman analysis detect a significant relationship [r=.001, N=193, *p*=.989].

5.3.1.2e *Context Provision strategies and vocabulary gain*

No significant relationship was detected between the variables of strategies which loaded onto the Context Provision factor and receptive vocabulary gain, either from parametric [r=.085, N=196, *p*=.234] or non-parametric [r=.069, N=196, *p*=.339] correlation analyses. Neither was a relationship found between the Context Provision strategies and productive vocabulary gain [r=.026, N=129, *p*=.767]. Moreover, there was no significant relationship between the Context Provision strategies and overall vocabulary gain as determined by both a Pearson [r=.040, N=193, *p*=.584] and a Spearman [r=.078, N=193, *p*=.280] analysis.

5.3.1.3 Summary of correlation analyses

There existed a significant but small, negative relationship between strategic behaviour and receptive and overall vocabulary gain, with higher levels of identification with strategic behaviour associated with lower receptive and overall vocabulary gain. No relationship was found between strategic behaviour and productive vocabulary gain.

In terms of the factored strategic behaviour data, there was a small, positive relationship between identification with strategies that loaded the Cognitive Action factor and receptive and overall vocabulary gain, indicating that higher identification with Cognitive Action strategies was linked to higher receptive and overall gain. There was also a relationship revealed between identification with the Establishment factor strategies and productive vocabulary gain: higher identification with strategies that loaded onto the Establishment factor was correlated with lower productive vocabulary gain. No other significant relationships were detected by the correlation analyses.

5.3.2 MULTIPLE REGRESSION ANALYSES

Hierarchical multiple regression analysis was used to determine the ability of the three independent variables (strategic behaviour, location and length of stay) to predict vocabulary gain. Regression was chosen as the analytic approach so that the independent variable of primary interest (identification with vocabulary-related strategic behaviour) could be represented on a continuous scale.

In this series of regression analyses strategic behaviour was represented by the five factor variables as opposed to the original strategic behaviour variable (which comprised the means of the responses to the strategic behaviour items in the questionnaire). This decision was taken because, as argued in section 2.3, strategies are a disparate and largely context-

specific set of behaviours and as such, to bunch them together would be to consider them conceptually homogenous and would therefore counter the theoretical standpoint. The five strategic behaviour factor variables were entered as block 1, and length of stay and location of SAE were entered together as block 2. The analyses took receptive, productive and overall vocabulary gain as the dependent variable respectively.

A supplementary regression model, in which the five factor variables were replaced in block 1 by the original strategic behaviour variable, is reported in Appendix M. To mirror the analyses described in 4.3.2, the original strategic behaviour variable was also dichotomised for use in a series of corresponding ANOVAs in order to investigate whether ANOVA would support the findings of the regression analyses. For the ANOVAs, the identification with strategic behaviour variable was transformed to make a binary nominal variable using a median split based on the sum of the mean of participants' reported strategic behaviour from the questionnaire; those with lower means rated the most number of statements as being very true of them and were thus labelled the High Identifiers (HI), and those with higher means were labelled the Low Identifiers (LI). The findings of these ANOVAs (provided in Appendix N) supported those of the regressions described in Appendix M.

5.3.2.1 Sample groupings

Strategic behaviour, represented in the model by the five factored variables, was taken as the independent variable of primary interest. In terms of the other two independent variables, the sample was divided into a three-level ordinal variable based on length of stay (group 1, 6-10 weeks; group 2, 11-15 weeks; group 3, 16-20 weeks) and into a binary categorical variable based on the location of the SAE (Oxford or London).

5.3.2.2 Parametric assumptions

Here follows a description of how the variables were investigated and treated to ensure they met the parametric assumptions required for regression analysis.

5.3.2.2a *Ratio of cases to independent variables*

As described in 4.3.2.2a, all three dependent variables (receptive, productive and overall vocabulary gain) exceeded the ratio of cases to independent variables advised by Tabachnick and Fidell (2001).

5.3.2.2b *Treatment of outliers*

A discussion of the treatment of outliers detected in the overall and receptive vocabulary gain variables is given in 4.1.1.4a and 4.1.1.4b respectively. An account of the logarithmic transformation of the productive vocabulary gain variable in response to outliers is provided in 4.1.1.4c. No outlaying cases existed in any of the independent variables.

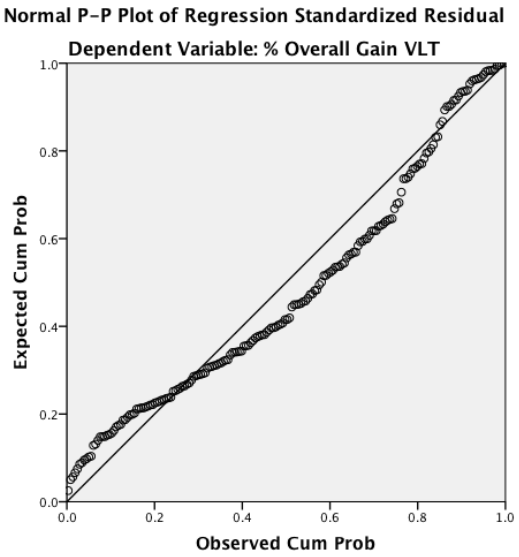
5.3.2.2c *Multicollinearity and singularity*

An initial regression was run to check the assumptions of multicollinearity, normality, linearity and homoscedasticity. The initial regression model took overall vocabulary gain as the dependent variable, the five factored strategic behaviour variables as block 1, and location and length of stay as block 2. The tolerance values for the independent variables as shown in the collinearity statistics output of the initial regression were all high (.539 to .893). Moreover, inspection of the correlations output revealed that none of the variables had a bivariate correlation higher than .35: Tabachnick and Fidell (1996) suggest that correlations above .7 are too high for inclusion of the variables in a regression model. Therefore, these figures provided strong evidence to suggest that the assumption of the absence of multicollinearity had not been violated.

5.3.2.2d *Normality, linearity and homoscedasticity of residuals*

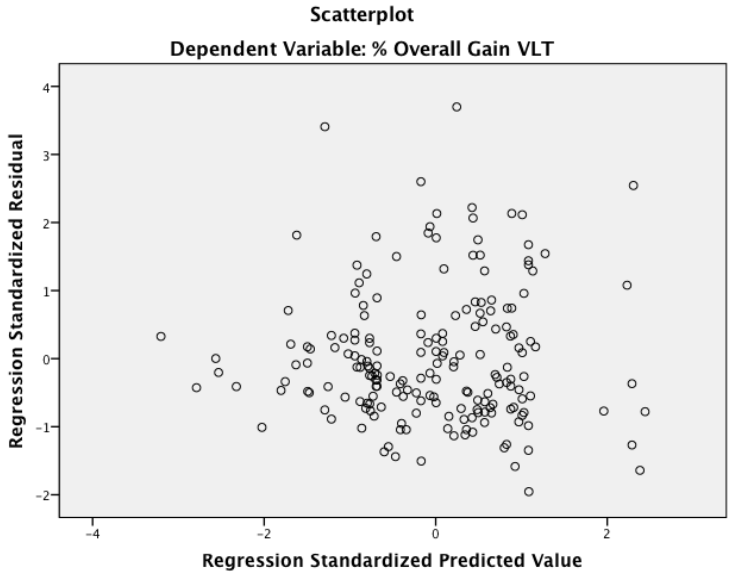
The normal probability plot (figure 5.8) exhibited a reasonably straight, diagonal linear manifestation of the points, which suggested that there was no major deviation from the assumption of normality.

Figure 5.8: Initial regression normal probability plot



Examination of the residuals scatterplot (figure 5.9) for the model indicated that there was very little deviation from normality and that heteroscedasticity and non-linearity of residuals were not an issue.

Figure 5.9: Initial regression residuals scatterplot



5.3.2.2e *Independence of errors*

The Durbin-Watson statistic of the final model of the initial regression was 2.01, suggesting that the residuals were uncorrelated and therefore that the assumption of independence of errors had not been violated.

5.3.2.2f *Solution outliers*

In order to inspect the model for solution outliers, the Mahalanobis distances produced by the initial regression were investigated. Tabachnick and Fidell (2001) state that the critical chi-square value for a model with eight independent variables is 26.13 and that any cases that exceed this value can be deemed as solution outliers. The model had eight independent variables because each of the five factored strategic behaviour variables and the two length of stay dummy variables were counted separately. No cases in the model had a Mahalanobis distance value exceeding 26.13: the minimum distance was 2.89 and the maximum 19.51 (M=7.96, SD=3.14).

5.3.2.3 Receptive vocabulary gain

A hierarchical multiple regression analysis was conducted to determine the extent to which the independent variables (block 1=the five factored strategic behaviour variables; block 2=length of stay and location of SAE) were able to predict the receptive vocabulary gain scores. After block 1 had been entered, the model explained 6.2% of the variance in receptive vocabulary gain. After the entry of block 2 the overall model explained 9.9%. Both model 1 [F(5, 189)=2.482, p=.033] and model 2 [F(8, 186)=2.55, p=.012] reached statistical significance. However, the R Square change value was non-significant (.037, p=.057), which suggested that when the effect of the factored strategic behaviour was controlled for, the variables in block 2 had no significant explanatory power.

Four of the independent variables made a statistically significant contribution to the overall model: identification with Appraisal strategies (β=.215, p=.017), identification with Visual Reference strategies (β=-.158, p=.041), identification with Establishment strategies (β=-.155, p=.046), and length of stay (β=-.239, p=.008). These findings suggest that stronger identification with Appraisal strategies predicted an increase in receptive vocabulary gain, whereas stronger identification with Visual Reference and Establishment strategies predicted a decrease in receptive gain. Moreover, belonging to the medium stay group predicted lower receptive gain.

Table 5.9: Predictor variable contribution to receptive gain regression with factored strategic behaviour

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	7.631	14.196		.538	.592
Cognitive Action	1.092	.440	.201	2.482	.014
Appraisal	.619	.515	.091	1.203	.231
Visual Reference	-1.414	.961	-.111	-1.471	.143
Establishment	-1.104	.571	-.147	-1.933	.055
Context Provision	.566	.860	.051	.658	.511
2 (Constant)	17.037	14.517		1.174	.242
Cognitive Action	.365	.514	.067	.709	.479
Appraisal	1.464	.605	.215	2.419	.017
Visual Reference	-2.014	.981	-.158	-2.054	.041
Establishment	-1.161	.578	-.155	-2.006	.046
Context Provision	.771	.874	.070	.883	.378
Location of SAE	-1.672	5.110	-.024	-.327	.744
Length of stay dummy variable 1: 6-10 wks?	2.085	6.425	.025	.325	.746
Length of stay dummy variable 2: 11-15 wks?	-19.542	7.322	-.239	-2.669	.008

5.3.2.4 Productive vocabulary gain

A hierarchical regression analysis was conducted to determine the extent to which the primary and secondary independent variables were able to predict the logarithmically transformed productive vocabulary gain variable. After block 1 had been entered, the model explained 4.5% of the variance in productive vocabulary gain but this finding was not significant [F(5, 123)=1.164, p=.331]. After the entry of block 2 the overall model accounted for 6.2% of the variance, but neither was this model significant [F(8, 128)=1.00, p=.440]. The R square change value was also non-significant (.017, p=.531).

Only one of the predictor variables made a significant contribution to the model: Stronger identification with Establishment strategies predicted lower productive vocabulary gain (β=-.219, p=.027).

Table 5.10: Predictor variable contribution to productive gain regression with factored strategic behaviour

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	1.657	.275		6.036	.000
Cognitive Action	.004	.009	.046	.455	.650
Appraisal	.005	.010	.052	.457	.586
Visual Reference	.002	.019	.012	.132	.895
Establishment	-.026	.011	-.227	-2.385	.019
Context Provision	.006	.017	.037	.376	.708
2 (Constant)	1.752	.285		6.145	.000
Cognitive Action	-.002	.010	-.019	-.159	.874
Appraisal	.011	.012	.108	.958	.340
Visual Reference	-.003	.019	-.024	-.244	.808
Establishment	-.025	.011	-.219	-2.237	.027
Context Provision	.008	.017	.049	.485	.628
Location of SAE	.004	.100	.004	.038	.970
Length of stay dummy variable 1: 6-10 wks?	-.074	.126	-.058	-.589	.557
Length of stay dummy variable 2: 11-15 wks?	-.205	.144	-.162	-1.423	.157

5.3.2.5 Overall vocabulary gain

Hierarchical multiple regression was also conducted to determine the extent to which the three independent variables were able to predict overall vocabulary gain. After block 1 had been entered, the model explained 5.8% of the variance in overall vocabulary gain. After block 2 had been entered, the model as a whole explained 9%. Both model 1 [F(5,186)=2.29, p=.048] and model 2 [F(8, 183)=2.26, p=.025] were statistically significant. The R square change value in the analysis was not significant (.032, p=.098), suggesting that the secondary independent variables failed to add any explanatory power to the model.

Two predictor variables made a significant contribution to the overall model: identification with Visual Reference strategies (β=-.2.03, p=.043) and length of stay (β=-.218, p=.017). The negative beta weight of the Visual Reference strategies variable in this model

suggested that a decrease in identification with the Visual Reference strategies was able to predict an increase in overall vocabulary gain. The negative beta weight reported with reference to length of stay indicated that membership of the medium stay group (11-15 weeks) predicted lower overall vocabulary gain.

Table 5.11: Predictor variable contribution to overall gain regression with factored strategic behaviour

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	11.404	11.142		1.024	.307
Cognitive Action	1.014	.345	.240	2.935	.004
Appraisal	.246	.404	.046	.608	.544
Visual Reference	-1.086	.754	-.110	-1.439	.152
Establishment	-.665	.448	-.114	-1.484	.140
Context Provision	-.056	.675	-.007	-.083	.954
2 (Constant)	17.545	11.429		1.535	.126
Cognitive Action	.618	.405	.147	1.526	.129
Appraisal	.678	.477	.128	1.422	.157
Visual Reference	-1.571	.772	-.159	-2.034	.043
Establishment	-.591	.455	-.101	-1.298	.196
Context Provision	.132	.688	.015	.192	.848
Location of SAE	-1.081	4.024	-.020	-.269	.788
Length of stay dummy variable 1: 6-10 wks?	-4.817	5.058	-.075	-.952	.342
Length of stay dummy variable 2: 11-15 wks?	-13.847	5.765	-.218	-2.402	.017

5.3.2.6 Summary of multiple regression analyses

The findings of the regression analyses here discussed indicate that the model was able to predict a small percentage of the variance in receptive and overall vocabulary gain. For receptive vocabulary gain, three of the strategic behaviour factors were significant contributors: Appraisal strategies predicted positive receptive gain whereas Visual Reference and Establishment strategies predicted negative gain. Additionally, the model found a significant contribution of length of stay, with belonging to the medium stay group a predictor of lower receptive gain.

For productive vocabulary gain, neither the initial nor the factored model was able to reveal any explanatory power. However, the factored model revealed a significant contribution of Establishment strategies, whereby stronger identification with these strategies predicted lower levels of productive gain.

For overall vocabulary gain, both strategic behaviour and length of stay made a significant impact. Once more, stronger identification with vocabulary-related strategies predicted a decrease in gain. Additionally, membership of the medium stay group predicted lower overall gain. When strategic behaviour was represented in the model by the factored strategic behaviour variables, identification with Visual Reference strategies and belonging to the medium stay group both predicted lower overall gain.

5.4 INFORMAL CONTACT, VOCABULARY-RELATED STRATEGIC BEHAVIOUR AND VOCABULARY GAIN

5.4.1 CORRELATION ANALYSES

Here follows the report of the parametric and non-parametric correlation analyses employed to determine whether there was a relationship between the independent variables of informal second language contact and strategic behaviour.

5.4.1.1 Informal contact and strategic behaviour

A Pearson product-moment analysis was conducted on the informal L2 contact and strategic behaviour variables, to find a medium-strength, positive, statistically significant relationship between the two variables [r=.301, N=240, *p*=.000]. Stronger identification with informal L2 contact scenarios was associated with stronger identification with vocabulary-related strategic behaviour. The coefficient of determination indicated that

identification with informal L2 contact scenarios explained 9% of the variance in the sample's identification with vocabulary-related strategic behaviour.

5.4.2 MULTIPLE REGRESSION ANALYSES

Hierarchical regression was employed to determine the ability of identification with informal L2 contact and identification with strategic behaviour, which were entered together as block 1, and length of stay and location of SAE, entered together as block 2, to predict the dependent variable of vocabulary gain. The regression analyses took receptive, productive and overall vocabulary gain as the dependent variable respectively. Informal L2 contact and strategic behaviour were represented in the model by the factor variables they begot.

A series of corresponding regression analyses, in which the factored informal contact variables were replaced by the original, mean scores variable is detailed in Appendix O. Additionally, a series of ANOVAs was conducted (in which both of the primary independent variables were dichotomised at the median split), the findings of which are detailed in Appendix P and support the regression findings described in Appendix O.

In this regression model the independent variables of primary interest were represented by the factor variables they gave rise to. In the case of informal L2 contact, EFA found three factors underlying the data. With regards strategic behaviour, five factors were found. A summary of the classifications given to the factors found is provided in table 5.12, in which the strongest weighted factors appear in the top row of the table and the weakest at the bottom.

Table 5.12: Factor variables of informal contact and strategic behaviour

Informal L2 Contact	Strategic Behaviour
Interactive	Cognitive Action
Literate	Appraisal
Narrative	Visual Reference
	Establishment
	Context Provision

The eight factor variables were entered into the model as block 1, and location and the length of stay dummy variables were entered as block 2. Receptive, productive and overall vocabulary gain were taken in turn as the dependent variable.

5.4.2.1 Parametric assumptions

Here follows a description of how the variables were investigated and treated to ensure they met the parametric assumptions required for regression analysis.

5.4.2.1a *Ratio of cases to independent variables*

As described in 4.3.2.2a, all three dependent variables (receptive, productive and overall vocabulary gain) exceeded the ratio of cases to independent variables advised by Tabachnick and Fidell (2001).

5.4.2.1b *Treatment of outliers*

A discussion of the treatment of outliers detected in the overall and receptive vocabulary gain variables is given in 4.1.1.4a and 4.1.1.4b respectively. An account of the logarithmic transformation of the productive vocabulary gain variable in response to outliers is provided in 4.1.1.4c. No outlaying cases existed in any of the independent variables.

5.4.2.1c *Multicollinearity and singularity*

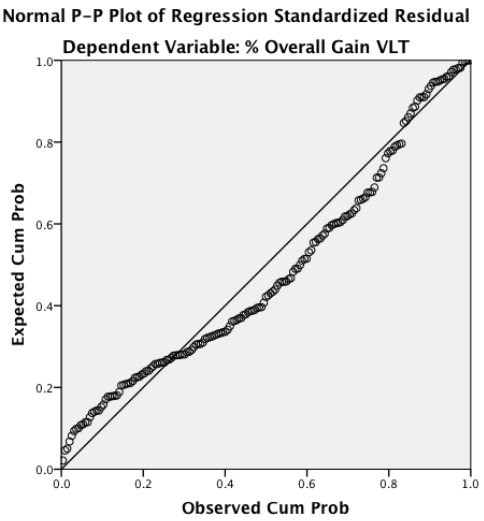
An initial regression was run to check the assumptions of singularity, normality, linearity and homoscedasticity. The initial regression model took overall vocabulary gain as the

dependent variable. The tolerance values for the eleven independent variables as shown in the collinearity statistics output of the initial regression model were all moderate to high (.436 to .829). These values provided evidence to suggest that the assumption of the absence of multicollinearity had not been violated.

5.4.2.1d *Normality, linearity and homoscedasticity of residuals*

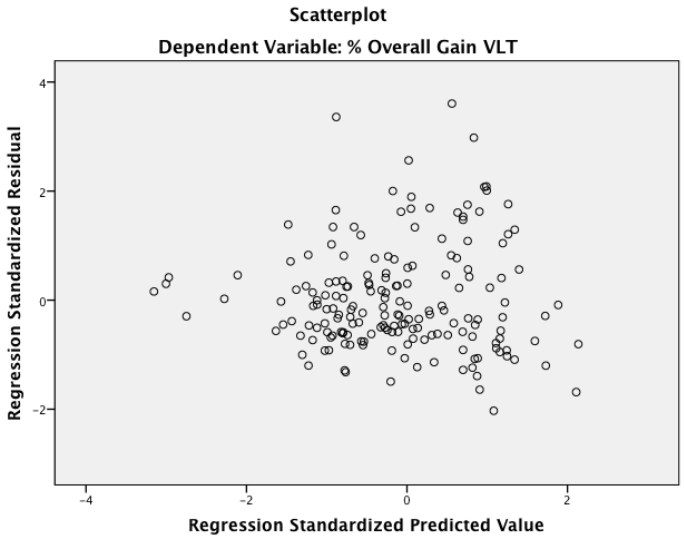
The normal probability plot generated from the initial model (figure 5.10) exhibited a predominantly straight, diagonal linear manifestation of the points, suggesting that the assumption of normality had not been violated.

Figure 5.10: Initial regression normal probability plot



The shape of the residuals scatterplot (figure 5.11) indicated that there was little deviation from normality and that heteroscedasticity and non-linearity of residuals were absent from the model.

Figure 5.11: Initial regression residuals scatterplot



5.4.2.1e *Independence of errors*

The Durbin-Watson statistic of the final model of the initial regression was 2.06, suggesting that the residuals were uncorrelated and that the assumption of independence of errors had not been violated.

5.4.2.1f *Solution outliers*

In order to inspect the model for solution outliers, the Mahalanobis distances produced by the initial regression were investigated. Tabachnick and Fidell (2001) state that the critical chi-square value for a model with eleven independent variables is 31.26 and that any cases that exceed this value can be deemed as solution outliers. No cases in the model had a Mahalanobis distance value exceeding 31.26: the minimum distance was 3.69 and the maximum 22.68 (M=10.93, SD=3.80).

5.4.2.2 Receptive vocabulary gain

A hierarchical multiple regression analysis was conducted to determine the extent to which the independent variables were able to predict receptive vocabulary gain scores. After the

entry of block 1 (the eight factored variables), the proportion of variance explained was 9% and identification with the factored informal contact and strategic behaviour variables combined was shown to have a significant effect on receptive vocabulary gain [$F(8, 176)=2.19$, $p=.031$]. After block 2 had been entered into the model, the model as a whole explained 14% of the variance and this model was also significant [$F(11, 173)=2.56$, $p=.005$]. The R Square change value was also significant (.049, $p=.021$), which suggested that when the effect of identification with the factored informal contact and strategic behaviour variables was controlled for, the variables in block 2 (location and length of stay) had significant explanatory power.

Seven of the eleven predictor variables in this model made a significant contribution: identification with Cognitive Action strategies ($\beta=.255$, $p=.034$), identification with Appraisal strategies ($\beta=.238$, $p=.016$), identification with Visual Reference strategies ($\beta=-.201$, $p=.013$), identification with Establishment strategies ($\beta=-.313$, $p=.002$), identification with Interactive scenarios ($\beta=-.206$, $p=.030$), identification with Narrative scenarios ($\beta=.259$, $p=.018$), and length of stay ($\beta=-.291$, $p=.002$).

The positive beta values pertaining to Cognitive Action and Appraisal strategies and Narrative scenarios indicated that stronger identification with these factors was predictive of receptive vocabulary gain. The negative beta values pertaining to Visual Reference and Establishment strategies and Interactive scenarios pointed to a decrease in receptive gain when identification with these factors increased. Moreover, belonging to the medium stay group was revealed to be predictive of lower receptive gain.

Table 5.13: Predictor variable contribution to final receptive gain regression equation

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	12.809	18.439		.692	.488
Cognitive Action	1.819	.590	.335	3.081	.002
Appraisal	.813	.601	.119	1.354	.178
Visual Reference	-1.705	1.006	-.134	-1.694	.092
Establishment	-2.048	.747	-.273	-2.742	.007
Context Provision	.034	.929	.003	.037	.971
Interactive scenarios	-.751	.387	-.170	-1.940	.054
Literate scenarios	-.229	.492	-.040	-.466	.642
Narrative scenarios	1.589	.888	.189	1.790	.075
2 (Constant)	27.012	18.748		1.441	.151
Cognitive Action	1.383	.649	.225	2.132	.034
Appraisal	1.625	.667	.238	2.436	.016
Visual Reference	-2.565	1.026	-.201	-2.499	.013
Establishment	-2.350	.752	-.313	-3.125	.002
Context Provision	.062	.938	.006	.066	.947
Interactive scenarios	-.913	.418	-.206	-2.185	.030
Literate scenarios	-.467	.492	-.081	-.949	.344
Narrative scenarios	2.177	.915	.259	2.380	.018
Location of SAE	-.561	5.352	-.008	-.105	.917
Length of stay dummy variable 1: 6-10 wks?	-4.881	7.142	-.059	-.683	.495
Length of stay dummy variable 2: 11-15 wks?	-23.862	7.620	-.291	-3.131	.002

5.4.2.3 Productive vocabulary gain

The regression analysis was then repeated, taking the logarithmically transformed productive vocabulary gain variable as the dependent variable. After the eight factor variables in block 1 had been entered, the overall model explained 7.7% of the variance in productive vocabulary gain and this finding was not statistically significant [F(8, 113)=1.18, *p*=.318]. After block 2 (location and length of stay) had been entered, the model as a whole explained 11.1% of the variance but again, this model did not reach statistical significance [F(11, 110)=1.25, *p*=.262]. Neither was the R square change statistic significant (.034, *p*=243).

Two of the predictor variables made a statistically significant contribution to the model: identification with Establishment strategies (β=-.284, p=.029) and identification with Interactive scenarios (β=-.265, p=.030). The negative beta values of these findings suggest that higher identification with these two factors served to predict lower productive vocabulary gain.

5.4.2.4 Overall vocabulary gain

Finally, the regression was also conducted to determine the extent to which the eleven independent variables were able to predict overall vocabulary gain. After block 1 had been entered, the model explained 7.4% of the variance but this model was not statistically significant [F(8,173)=1.72, p=.096]. After block 2 had been entered, the model as a whole explained 12.4% of the variance and this model did reach significance [F(11,170)=2.18, p=.017]. Moreover, the R square change value in this regression analysis was significant (.050, p=.024), suggesting that the secondary independent variables contributed explanatory power to the model when the primary independent variables had been controlled for.

Five of the eleven predictor variables made a significant contribution to the overall model: identification with Cognitive Action strategies (β=.330, p=.007), identification with Visual Reference strategies (β=-.195, p=.018), identification with Establishment strategies (β=-.259, p=.012), identification with Narrative scenarios (β=.253, p=.023), and length of stay (β=-.271, p=.005).

The positive beta weight pertaining to Cognitive Action and Visual Reference strategies and Narrative scenarios suggested that stronger identification with these factors predicted higher overall vocabulary gain. The negative beta values with reference to Establishment strategies and length of stay indicated that stronger identification with Establishment

strategies and belonging to the medium stay group were both predictive of lower overall vocabulary gain.

Table 5.14: Predictor variable contribution to final overall gain regression equation

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	13.865	14.584		.951	.343
Cognitive Action	1.519	.467	.360	3.254	.001
Appraisal	.217	.475	.041	.457	.648
Visual Reference	-1.250	.796	-.126	-1.570	.118
Establishment	-1.296	.591	-.222	-2.193	.030
Context Provision	-.419	.735	-.049	-.570	.569
Interactive scenarios	-.283	.306	-.082	-.924	.357
Literate scenarios	-.322	.389	-.072	-.827	.409
Narrative scenarios	1.126	.702	.172	1.603	.111
2 (Constant)	25.986	14.834		1.752	.082
Cognitive Action	1.393	.513	.330	2.713	.007
Appraisal	.716	.528	.135	1.356	.177
Visual Reference	-1.932	.812	-.195	-2.379	.018
Establishment	-1.507	.595	-.259	-2.533	.012
Context Provision	-.444	.742	-.052	-.599	.550
Interactive scenarios	-.528	.331	-.153	-1.596	.112
Literate scenarios	-.529	.389	-.118	-1.358	.176
Narrative scenarios	1.656	.724	.253	2.288	.023
Location of SAE	.107	4.235	.002	.025	.980
Length of stay dummy variable 1: 6-10 wks?	-9.221	5.651	-.144	-1.632	.105
Length of stay dummy variable 2: 11-15 wks?	-17.244	6.029	-.271	-2.860	.005

5.4.2.5 Summary of multiple regression analyses

The findings of the regression analyses here discussed indicate that the model was able to explain 14% of the variance in receptive vocabulary gain and 12.4% of the variance in overall gain. For receptive gain, stronger identification with Cognitive Action and Appraisal strategies and Narrative scenarios predicted higher gain. Stronger identification with Visual Reference and Establishment strategies and Interactive scenarios, and belonging to the medium stay group predicted lower gain. In terms of overall vocabulary

gain, identification with Cognitive Action and Visual Reference strategies and Narrative scenarios had a positive relationship with gain, whereas identification with Establishment strategies and belonging to the medium stay group predicted lower levels of gain. The model had no explanatory power over productive vocabulary gain.

5.5 POST-HOC STATISTICAL ANALYSES

5.5.1 VOCABULARY GAIN BY LENGTH OF STAY

In section 4.1.3.2a the receptive, productive and overall vocabulary gain variables were taken as the dependent variable respectively in a series of one-way, between groups ANOVAs to determine the impact of length of stay. The ANOVAs found a statistically significant difference between the three length of stay groups on receptive vocabulary gain [$F(2, 193)=3.05$, $p=.05$] with a small effect size ($\eta^2=.03$) and overall vocabulary gain [$F(2, 190)=3.75$, $p=.025$] with a small effect size ($\eta^2=.04$). There was no difference between the groups on productive vocabulary gain.

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 percentage 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 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 (η^2=.18), yet for the long stay group (p=.000) the effect size for overall vocabulary gain was large (η^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 the SAE on the overall vocabulary gain of the long stay group was large.

5.5.2 IDENTIFICATION WITH INFORMAL L2 CONTACT BY LOCATION

In section 4.2.3.2a an independent-samples *t*-test was conducted on the mean scores of the individual informal L2 contact items from the questionnaire with the sample divided into two groups based on location. The results showed that in response to 11 of the 28 informal contact scenarios, the Oxford and London groups' means were significantly different from one another. A Bonferroni adjustment (0.05 / 28) was applied to find that only three of the 11 differences reached significance, achieved at the p<.002 level. The significant differences were:

1. Listen to English TV/radio/podcasts
2. Read subtitles in English whilst watching TV/movies
3. Overall writing

Why might learners on a SAE in Oxford identify more highly with scenarios involving watching television 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 TV and writing in the L2.

5.5.3 IDENTIFICATION WITH STRATEGIC BEHAVIOUR BY LOCATION

In section 5.2.4.1a an independent-samples *t*-test was conducted to determine whether there was a difference between the participants as grouped by location (Oxford and London) on the means of their responses to the strategic behaviour questionnaire items. There was a significant difference in identification with vocabulary-related strategic behaviour between participants in Oxford (N=129, M=2.62, SD=.48) and participants in London (N=112, M=2.77, SD=.53; $t(239)=-2.32$, $p=.02$). This result was mirrored by the significant finding of a corresponding Mann-Whitney U test [$z=-2.68$, $p=.007$). As the mean value of the Oxford group for this variable was lower than that of their counterparts in London, this finding indicated that participants in Oxford identified more highly with the vocabulary-related strategic behaviour (because 1=This is very true of me and 5=This is not at all true of me). The magnitude of the differences discerned by the *t*-test was small ($\eta^2=.022$).

Why might the participants in Oxford have identified with the vocabulary-related strategies more highly than the London-based group? An independent-samples *t*-test further revealed that there was a significant difference between the mean proficiency level of the participants in Oxford (M=2.81, SD=.87) and the participants in London [M=2.58, SD=.89; $t(239)=2.06$, $p=.041$]. The magnitude of the differences was small ($\eta^2=.02$). Therefore, the participants whose proficiency level was higher identified more highly with vocabulary-related strategic behaviour.

5.5.4 IDENTIFICATION WITH STRATEGIC BEHAVIOUR BY LENGTH OF STAY

In section 5.2.4.2a a one-way between-groups analysis of variance was conducted, taking the mean responses to the 28 strategic behaviour items from part 2 of the questionnaire as the dependent variable. There was a statistically significant difference in identification with vocabulary-related strategic behaviour for the three groups [F(2, 238)=4.92, *p*=.008). A corresponding Kruskal-Wallis test also found a significant difference between the sample identification with strategic behaviour as grouped by length of stay ($\chi^2(2)=9.33$, *p*=.009). Post-hoc ANOVA comparisons using the Tukey HSD test indicated that the mean scores on identification with strategic behaviour for the short stay group (M=2.81, SD=.40) were significantly different from those of the medium stay group (M=2.52, SD=.44; *p*=.007). The long stay group (M=2.71, SD=.56) did not significantly differ from the short stay group but differed significantly from the medium stay group (*p*=.047). According to these findings, participants in the medium stay group identified more strongly with the strategic behaviour items in the questionnaire than did participants in the short stay and long stay groups (because 1=This is very true of me and 5=This is not at all true of me).

In light of this finding, a post-hoc ANOVA on proficiency level by length of stay was conducted to find that there was a significant difference between the length of stay groups [F(2, 238)=3.9, *p*=.022]. Post-hoc comparisons using the Tukey HSD test indicated that the mean proficiency level of the medium stay group (M=2.98, SD=1.08) was significantly higher (*p*=.016) than that of the long stay group (M=2.59, SD=.88). The group with the highest proficiency level was also the group that identified most highly with vocabulary-related strategies.

5.6 SUMMARY OF QUANTITATIVE STRATEGIC BEHAVIOUR

The top ten most identified with strategies were:

1. Use a bilingual dictionary to find the L1 translation of the word/phrase
2. Repeat the word/phrase in my head to try to remember it
3. Try to think of a translation of the word/phrase in my native language
4. Ask the speaker to explain (in English) the meaning of the new word/phrase
5. Try to guess the meaning from looking at the word
6. Try to remember the word/phrase to ask teacher/friends later for the meaning
7. Write down the word/phrase and its meaning immediately
8. Try to decide the best way to learn the word/phrase
9. Try to connect the new word/phrase to a personal experience I have had
10. Try to think of situations where I could use the new word/phrase

There was a significant difference in identification with vocabulary-related strategic behaviour between participants undertaking their SAE in Oxford and in London, with participants in Oxford identifying more highly with the strategies from part 2 of the questionnaire than their counterparts in the capital. A post-hoc independent-samples *t*-test further revealed that there was a significant difference between the higher mean proficiency level of the participants in Oxford than the participants in London. Therefore, the location group that had a higher proficiency level identified more strongly with vocabulary-related strategic behaviour.

In terms of length of stay, the medium stay participants identified significantly more strongly with the strategic behaviour items than the other two groups. A post-hoc ANOVA on proficiency level by length of stay found that there was a significant difference between the length of stay groups, which indicated that the mean proficiency level of the medium

stay group was significantly higher than that of the long stay group. As such, the length of stay group with the highest proficiency level was also the group that identified most highly with vocabulary-related strategies.

Factor analysis revealed five underlying groups of vocabulary-related strategies: Cognitive Action strategies, comprising strategies that were solitary, solely cognitive and comprise preparation for using the lexical item; Appraisal strategies, which were concerned with judging an item in terms of what it means and if/how best to learn it; Visual Reference strategies, which make use of visual information about the item in order better to determine its meaning and use; Establishment strategies, which would most likely be deployed soon after first contact with an unknown item and which involved establishing a basis from which to put the lexical item to work; and Context Provision strategies, whereby the learner decided upon the linguistic contexts in which the lexical item could be used and how they could fit the item into their existing schemata. The factor that contained the highest number of strategies and onto which the strategies were most heavily weighted was the Cognitive Action factor.

The regression analyses of factored strategic behaviour and vocabulary gain indicated that the model was able to predict a small percentage of the variance in receptive and overall vocabulary gain. For receptive vocabulary gain, four factors were significant contributors: Appraisal strategies predicted more gain whereas Visual Reference and Establishment strategies and belonging to the medium stay group predicted negative gain. For overall vocabulary gain, both strategic behaviour and length of stay made a significant impact. Identification with Visual Reference strategies and belonging to the medium stay group both predicted lower overall gain.

5.7 SUMMARY OF OVERALL FINDINGS

In this section is detailed a summary of the findings with reference to the three research questions.

5.7.1. RESEARCH QUESTION 1

Is there a relationship between identification with informal L2 contact scenarios and vocabulary gain in a study abroad context?

The findings of the correlation and multiple regression analyses in this study found no significant relationship between identification with informal L2 contact scenarios and vocabulary gain.

5.7.2 RESEARCH QUESTION 2

What vocabulary-related strategies do learners of English employ in informal L2 contact in a study abroad context?

The qualitative data yielded by the OWLS instrument in semi-structured interviews revealed that the strategies employed by learners of English in informal L2 contact in a study abroad context are heavily dependent on the learner, the context, and the target lexical item. Factor analysis of the questionnaire data revealed five underlying groups of vocabulary-related strategies, of which the Cognitive Action factor, which comprised strategies that were solitary, solely cognitive and involved using or preparing to use the lexical item, contained the highest number of strategies. Moreover, the strategies that loaded onto the Cognitive Action factor weighed onto it more heavily than did the strategies in the other four factors.

5.7.3 RESEARCH QUESTION 3

Is there a relationship between vocabulary-related strategy use in informal L2 contact and vocabulary gain in a study abroad context?

The findings of the regression analyses indicated that for receptive vocabulary gain, stronger identification with Cognitive Action and Appraisal strategies and Narrative scenarios predicted higher gain. Stronger identification with Visual Reference and Establishment strategies and Interactive scenarios, and belonging to the medium stay group were found to predict lower gain. The model had no explanatory power over productive vocabulary gain. In terms of overall vocabulary gain, identification with Cognitive Action and Visual Reference strategies and Narrative scenarios had a positive relationship with gain, whereas identification with Establishment strategies and belonging to the medium stay group once more predicted lower levels of gain.

CHAPTER VI

DISCUSSION

In Chapter II the detailed analysis of the literature provided a comprehensive review of the current evidence on the relationships between informal second language contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context, to reveal gaps in the collective knowledge; gaps which the present study has sought to address. Firstly, the research into informal L2 contact has tended toward either (1) descriptive case studies that focused on very small numbers of participants, or (2) larger scale studies that required participants retrospectively to state the frequency with which they had encountered certain L2 contact scenarios. As such, there is a dearth of reliable L2 contact data from which generalisations can be made. Secondly, research into vocabulary-related strategic behaviour that occurs outside the boundaries of the laboratory or classroom has without exception required respondents to generalise their behaviour across contexts and/or tasks. As a result, the evidence to date on vocabulary-related strategy use in informal L2 settings must be considered cautiously in terms of its reliability and validity. Finally, no present evidence exists on the interplay between informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context.

The present study has addressed the issues that emerged from the literature review by contributing to the body of knowledge in three ways. Firstly, by investigating informal L2 contact in a study abroad context quantitatively at the temporal point at which that contact was taking place and from the point of view of identification with L2 contact scenarios as opposed to recollection of their frequency. Secondly, through the generation of thick, descriptive, context and task specific data on vocabulary-related strategic behaviour that

occurs in informal settings in tandem with numeric data on vocabulary-related strategy use which is specific to informal contact and did not force respondents to generalise across contexts. Finally, by examining the interplay between informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context to provide insight where previously there was none. This chapter provides a detailed discussion of the findings of the present study with reference to these three contributions to the body of knowledge of SLA research.

6.1 INFORMAL SECOND LANGUAGE CONTACT IN A STUDY ABROAD CONTEXT

Is there a relationship between identification with informal L2 contact scenarios and vocabulary gain in a study abroad context?

Identification with informal L2 contact was measured using a 5-point 'how true of me scale' in tandem with the 28 items in part 1 of the questionnaire. The order in which all 28 informal contact scenarios were identified with by the sample is detailed in 4.2.3. The top ten most highly identified with informal L2 contact scenarios, in rank order from most to least identified with, are shown in table 6.1, adjacent to the top ten most identified with informal L2 contact scenarios as found in the only other study which has attempted to determine the specific L2 contact scenarios most identified with by SAE learners (Briggs, 2014).

Table 6.1: Identification with informal contact: Present study and Briggs (2014)

Rank	Present study	Briggs (2014)
1	I speak in English to service personnel	I speak in English to service personnel
2	I use English to get information from people I don't know	I read timetables, announcements, posters, menus etc. in English outside of class
3	I listen to English songs outside of class	I listen to English songs outside of class
4	I read timetables, announcements, posters, menus etc. in English outside of class	I use English for short exchanges with the people I live with
5	I use English for short exchanges with the people I live with	I use English to get information from people I don't know
6	I listen to other people's conversations in English outside of class	I listen to English movies and videos outside of class
7	I listen to English movies and videos outside of class	I read emails/webpages in English
8	I read emails/webpages in English	I listen to other people's conversations in English outside of class
9	I speak English to the people I live with	Overall, I think I listen to a lot of English outside of class
10	In England, outside of class, I speak to my teacher/instructor	I speak English to the people I live with

As demonstrated in table 6.1, nine of the ten informal contact scenarios most identified with by the sample of the present study were also in the top ten most identified with scenarios of Briggs (2014). Moreover, the order in which these nine scenarios were identified with by the samples of the two studies is very similar. In both studies there was an approximately equal spread of interactive and solitary scenarios. It is interesting to note that of the interactive scenarios, none involved speaking to interlocutors outside of the home or school network: this finding suggests that SAE learners who live alone or with people who share their L1 are consequently deprived of a key opportunity for informal contact with the L2. In other words, if you do not live with L2 native speakers during your SAE, it is likely that you will miss out on the majority of the most highly identified with interactive informal L2 contact scenarios.

Factor analysis on the informal contact data from the present study revealed three underlying groupings: Interactive scenarios, which pertained to speaking and listening to the L2 in informal settings; Literate scenarios, which involved reading and writing in the

L2; and Narrative scenarios, which comprised understanding or communicating stories of some kind. The factor that contained the highest number of scenarios and onto which the scenarios were most heavily weighted was the Interactive factor. Of the top ten most identified scenarios, five loaded onto the Interactive (speaking and listening) factor (numbers 3, 5, 6, 9 and 10), one (number 4) loaded onto the Literate (reading and writing) factor, and one (number 7) onto the Narrative (comprehending/producing narratives) factor. As such, the informal L2 contact scenarios most identified with by the sample were Interactive scenarios.

In Briggs (2014) factor analysis similarly revealed three underlying informal contact factors. They were: (1) Individual scenarios, which all involved solitary contact with the L2; (2) Interactive, of which the vast majority of scenarios involved speaking or listening to the L2; and (3) Informative scenarios, which all involved the obtainment of information for practical or enjoyment purposes. The informal contact scenarios most highly identified with by the 2014 sample loaded onto the Informative factor, which comprised six scenarios.

There are certainly links to be drawn between the findings related to informal L2 contact of Briggs (2014) and the present study. Firstly, both studies found three informal contact factors. The 2014 study sampled participants only from Oxford, whereas the participants in the present study were sampled from Oxford and London. That the responses of both samples to the informal contact items manifested a triumvirate factor structure, and given that evidence was pursued to support the match between sample and population (c.f. 3.4), suggests that a similarly structured factor may be present in the informal L2 contact of study abroad learners across England. The implications of this finding, discussed in more detail in 7.3 and 7.4, are broad, particularly in terms of preparing learners for the kinds of informal contact to expect on a SAE and how strategically to manipulate that contact, and

in terms of researching the relationships between language contact, strategic behaviour and language acquisition.

Not only did both the present study and Briggs (2014) find a triumvirate informal L2 contact factor structure, but both found very similar factors within that structure. Both found an Interactive (speaking and listening) factor, and this suggests that interactive scenarios are a pervasive and consistent feature of informal L2 contact, irrespective of location of the SAE. The Literate (reading and writing) factor from the present study and the Individual (solitary contact) factor from the 2014 study are very similar in that the majority of scenarios that loaded onto them involve solitary contact with the L2. It appears to be common sense that both interactive and individual scenarios would feature highly in terms of the kinds of informal L2 contact that SAE learners are experiencing. However, the fact that these groupings arose from the data itself suggests that the differentiation between these two categories of contact exists within the minds of the respondents, and as such there is evidence to suggest that pedagogical approaches to preparing learners for SAE should take account of this division.

Five of the six Informative (information for practical or enjoyment purposes) scenarios from Briggs (2014) derived from the top ten most identified with scenarios in that study, and four of the six also appeared in the top ten most identified with scenarios in the present study. The six Informative scenarios are listed below and those that feature in the top ten of both the 2014 and the present study are marked with a double asterisk:

1. I read emails/web pages in English outside of class**
2. I use English to get information from people I don't know**
3. Overall, I think I read a lot in English outside of class

4. I read timetables, announcements, posters, menus etc. in English outside of class**
5. I read novels in English outside of class
6. I listen to English movies and videos outside of class**

The informal L2 contact scenarios that appeared in the top ten of both studies reveal that learner-initiated obtainment of information from a variety of sources (online, 'real', interpersonal, auditory, visual etc.) pervades the kinds of informal L2 contact that study abroad learners identify with. To summarise, the weight of the evidence presented in both Briggs (2014) and the present study suggests that there is a triumvirate structure to informal L2 contact in SAE, that within that structure there is a division between social and solitary contact with the L2, and that across the structure the most highly identified with scenarios involve self-initiated obtainment of information from external sources.

The short stay participants identified more strongly with the informal L2 contact scenarios than did the medium or long stay group. This is an interesting finding because one might assume that the short stay group had experienced the least exposure to informal L2 contact. However, it 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. There was no significant difference between the location groups based on the means of their responses to the informal contact items overall. However, the location groups did respond differently to three individual informal contact items. The scenarios with which they identified differently were:

1. I listen to English TV/spoken radio programmes/podcasts outside of class.
2. I read subtitles while I am watching TV or movies in English outside of class.
3. Overall, I think I do a lot of writing in English outside of class.

In the case of the first and second listed scenarios, the participants in Oxford identified more highly. In the case of the third, the London-based participants identified with the scenario more highly. A clear link can be drawn between the first two scenarios in that they both pertain to contact with English through watching television.

Why might learners on a SAE in Oxford identify more highly with scenarios involving watching television than their counterparts in the capital? In investigating this question a chi-square test of independence (c.f. 5.5.2) 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. As such, one hypothesis for the stronger identification by Oxford-based participants with scenarios involving watching television may be that the Oxford participants 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. It may also be the case that differentiation between the location groups in terms of television watching habits during the SAE accounted for their differential identification with the third scenario; overall writing. That is to say, it is possible that participants who chose not to watch or did not have access to English-language programmes instead engaged in more or a wider variety of written L2 contact (e.g. instant messaging, forums, chat rooms, emails, exam preparation).

6.2 VOCABULARY-RELATED STRATEGIC BEHAVIOUR IN INFORMAL SECOND LANGUAGE CONTACT IN A STUDY ABROAD CONTEXT

What vocabulary-related strategies do learners of English employ in informal L2 contact in a study abroad context?

6.2.1 QUALITATIVE STRATEGIC BEHAVIOUR

The qualitative findings with reference to vocabulary-related strategic behaviour are now discussed with reference to (1) the quantity of strategic responses to the OWLS scenarios and (2) the types of strategic responses revealed in the OWLS interview data.

6.2.1.1 Quantity of strategic responses

The most commonly reported classification of strategy in response to the OWLS scenarios was individual determination (working out the meaning of the word by themselves). This is perhaps unsurprising given that many Individual Determination strategies are available to the learner regardless of the physical features of the context in which a new word/phrase is encountered. However, surprisingly small numbers of individual Memorisation and Metacognition strategies were reported, although these strategy classifications did feature regularly in the reported chains and clusters. One possible reason for this finding is that the sample for the most part were concerned firstly with determining meaning (hence the preponderance of reported Individual Determination strategies), and thus Memorisation was less likely to be reported independently from a determination strategy. Similarly, as Metacognition involves the orchestration of cognitive strategies it was more likely that metacognitive strategies would be reported as part of a combined strategic response.

The OWLS scenario that yielded the highest overall number of strategic responses was simulation 5, which depicted listening to a song in English. As demonstrated in table 6.1, this scenario featured in the top three most identified with by the total sample of the present study and by the sample of Briggs (2014). Simulation 5 also prompted the highest number of combinations of strategies, and in particular the manifestation of strategic chains. One hypothesis for the findings with reference to simulation 5 is that, of the top ten most identified with scenarios, listening to songs in English is the scenario that offers a

language learner the most control and the greatest opportunity to test and develop their strategic responses: To actively be listening to music suggests that the learner is alone, and thus the gamut of determination strategies, excepting those which involve an interlocutor, can be employed without fear of disrupting interactional flow. Moreover, if the learner is listening on a laptop, smartphone or mp3 device, they can pause, rewind and repeat any part of a song they hear. As such, a number of Memorisation strategies (e.g. making a note of the word/phrase, repeating the stimulus) are available that are prohibited by continuous, 'online' exposure to a stimulus (e.g. as part of a TV programme or film). The control that learners are able to exert over this scenario allows for repeated practice of their strategic responses therein, and I contend that it is this aspect of simulation 5 that gave rise to the preponderance of chained responses. Also, the prevalence of websites that list the full lyrics of popular songs means that the need for a learner to have heard a target item sufficiently to guess its orthographic form (in order to look it up in a dictionary/online) is bypassed. In fact, the OWLS data suggests that strategy deployment in response to listening to English songs is more likely prompted by an appreciation of the music than by the presence of an unknown lexical item. Finally, the affective impact of listening to music may serve as a motivational factor in this scenario; spurring the manifestation of strategies to enhance comprehension of the lyrics and thus consolidating the learner's emotional connection to the music.

The simulation that yielded the fewest strategic responses was simulation 8, which depicted a roadside billboard advertisement. Almost all of the strategic responses to simulation 8 were instances of individual strategy use, and of those instances the vast majority were reports of Individual Determination strategies. Conversely to simulation 5, in real exposure to the scenario in simulation 8 the learner's contextual intention would unlikely ever be to acquire language; the contextual intention would most likely be related

to driving, such as how to get to where they are going or what they plan to do when they reach their destination. Moreover, a learner exerts no control over what they encounter as they are driving: unlike with simulation 5, they cannot prepare for or practice their strategic responses in this scenario without disrupting their central activity in it, which is driving. Clearly, the physical nature of driving inhibits a range of possible vocabulary-related strategies (such a dictionary reference or noting down the item and its meaning), and even if driving itself is automatic for a learner, the attention to the road required whilst driving may inhibit a range of solely cognitive strategies too. Moreover, the context in which the target word was presented in simulation 8 (i.e. as part of an advertisement) appears to have precluded the employment of a greater number of strategies in response to it: that is, in many cases, once the participant had recognised the billboard as an advertisement, the comprehension of the individual item was overshadowed by the application of background knowledge to determine that the message was encouraging the consumer to buy that particular chocolate bar.

6.2.1.2 Types of strategic responses

The findings from the OWLS interview data on the types of strategies and strategic combinations manifested in response to the simulated scenarios provide robust evidence in support of the context- and task-specific nature of strategic behaviour. Many researchers have argued for the vital roles of context and/or task in strategy use (Chamot, 2004; Macaro, 2004; Takeuchi, Griffiths & Coyle, 2007; Briggs, 2014) and the strength of the influence of these factors is clearly demonstrated in the present OWLS data. As in Briggs (2014), the data demonstrates examples of physical context prohibiting certain strategic behaviours; for example, in simulation 8 (read timetables, announcements, posters, menus etc.) the participant is driving along a motorway and this physical consideration is likely to account for the dearth of dictionary-related strategic behaviour reported in response to the

target lexical item in the simulation. Moreover, certain non-physical contextual features appear to have impacted on strategic action; for example, the sound of the telephone ringing in simulation 7 (short exchanges with cohabitants) and the urgent physical action that the sound prompts (i.e. there is only a short amount of time in which to answer it) may have accounted for the very small number of strategy combinations in response to the target lexical item. That is, the participants limited their available processing and strategic action time in order to be able to answer the telephone before the call was abandoned.

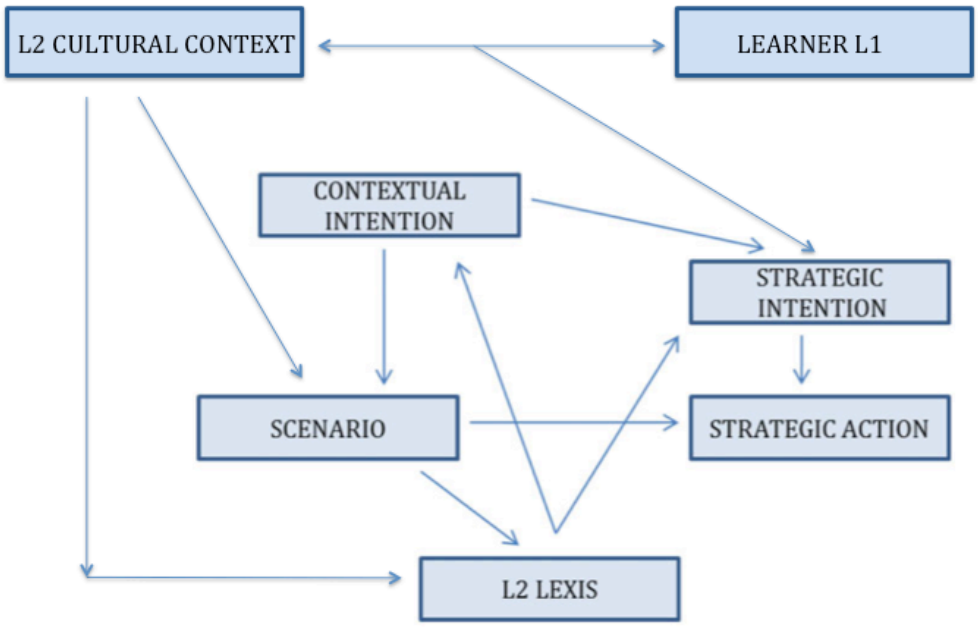
It appears from the OWLS data that the respondents made certain assumptions about the simulated scenes therein which in turn limited the types of available strategies from which they had to draw. For example, scenario 1 (speak English with cohabitants) depicted a group of friends seated around a dining table – there were at least six people at the table and they comprised a visually obvious diverse range of ethnicities – and yet in response to the target item 'fellow' not one deployed any strategy that involved a speaker of their L1. This finding therefore suggests that the sample assumed that the people sitting around the table were all native speakers of English who could not speak their L1. Similarly in response to simulation 4 (speak English to service personnel), the participants did not use any strategies that utilised the L1 and thus appeared to have assumed that the shop assistant in the simulation was a monolingual speaker of English. It is conceivable that, for example, a Japanese learner of English in England would assume that most native speakers of English would not be able to speak their L1, but it was surprising that the French, German and Spanish OWLS interviewees seemed to make the same assumption. Moreover, in simulation 5 (listen to songs in English), no interactive strategies were reported, and thus one could argue that the sample assumed that the woman who was listening to music was alone.

The apparent assumptions made by the sample about the OWLS scenarios beg the

question, ‘Would a different population in a different SAE context have made similar assumptions?’ I contend not. Just as there is evidence to suggest that a language learner’s culture impacts upon their strategy use (O’Malley & Chamot, 1990; Olivares-Cuhat, 2002; Keatley, Chamot, Spokane & Greenstreet, 2004), so too could the relationship between the L1 and the perceived culture of the target language community in a SAE have an impact. For example, a L1 English learner of French as a second language who is on a SAE in Paris is more likely to assume that any potential interlocutor is able to speak their L1 to a lesser or greater extent than is, say, a L1 Thai learner of L2 English in London. This is because the perception of the L2 cultural context by the SAE learner includes the belief that most native speakers of the L2 will also speak their L1. This means that the L1 English learner of French in France may assume a wider variety of potential vocabulary-related strategies because they have available to them the full range of interactive strategies (i.e. those which use the L2 and those which make use of the L1). This finding provides further evidence in support of the highly context and learner-specific nature of strategy use.

Briggs (2014) posited a model of the relationships between contextual intention (the learner’s overarching reason for engaging in a scenario), strategic intention (the learner’s reason for engaging in strategic behaviour), strategic action (the actual cognitive and/or physical strategic action the learner manifests) and the unknown lexis that a learner might encounter in a given informal L2 contact scenario. This model has hereby been extended to incorporate the impact of the relationship between the perceived cultural context of L2 and the learner’s L1 on the strategic experience of informal contact that a SAE learner may have.

Figure 6.1: Extended model of influencing factors in informal L2 contact



The model in figure 6.1 can further be exemplified by extending Macaro's (2004:4) algorithm for learner strategy application, which reads: "IF in a learning situation X, AND when the learning goal is Y, THEN try mental action Z". An extension of this algorithm that accounts for the L1 of the learner in relation to the cultural context of the SAE, the contextual intention of the learner, their strategic intention, and the lexis they may encounter reads: IF L1 is T, AND when in cultural context U, AND in scenario V, AND when contextual intention is W, IF meets lexical item X, AND strategic intention is Y, THEN try strategic action Z. For example, IF you are a L1 speaker of Spanish, AND you are learning English in the UK, AND you are having dinner with your host family, AND you want to use this opportunity to try to acquire some new L2 vocabulary, AND your host mother says the word 'vain', AND you like the sound of this word and want to understand and be able to use the word in future, THEN try asking your interlocutor for an explanation of the meaning in English and imagine how it is spelled.

6.2.2 QUANTITATIVE STRATEGIC BEHAVIOUR

Identification with vocabulary-related strategic behaviour was measured quantitatively using a 5-point 'how true of me scale' in tandem with the 28 items in part 2 of the questionnaire. The order in which all 28 of the vocabulary-related strategies were identified with by the sample is detailed 5.2.3. The top ten most identified with strategies were:

1. Use a bilingual dictionary to find the L1 translation of the word/phrase
2. Repeat the word/phrase in my head to try to remember it
3. Try to think of a translation of the word/phrase in my native language
4. Ask the speaker to explain (in English) the meaning of the new word/phrase
5. Try to guess the meaning from looking at the word
6. Try to remember the word/phrase to ask teacher/friends later for the meaning
7. Write down the word/phrase and its meaning immediately
8. Try to decide the best way to learn the word/phrase
9. Try to connect the new word/phrase to a personal experience I have had
10. Try to think of situations where I could use the new word/phrase

The top seven of these strategies are all related to very basic word-meaning repetition/connection: none involve use of the context (textual or background) in which the lexical item was encountered or analysis of the orthographic, phonological or grammatical features of the lexical item. In this way these findings mirror those of Lawson and Hogben (1996), whose observation of n=15 L2 learners of Italian revealed that the most frequently reported strategies were those that did not involve 'complex elaboration' with regard to the context in which the lexical item was encountered. However, the similarity between these findings is less robust in consideration of the fact that strategy use is heavily dependent

upon task and context. That is to say, Lawson and Hogben (1996) interviewed their participants in laboratory style, explicit vocabulary learning sessions in an EFL setting and the participants in the present study completed the questionnaire about strategic behaviour in informal L2 contact in a SAE setting, so the context and task were both different in the two studies.

Factor analysis revealed five underlying groups of vocabulary-related strategies: Cognitive Action strategies, comprising strategies that were solitary, solely cognitive and comprise preparation for using the lexical item; Appraisal strategies, which were concerned with judging an item in terms of what it means and if/how best to learn it; Visual Reference strategies, which make use of visual information about the item in order better to determine its meaning and use; Establishment strategies, which would most likely be deployed soon after first contact with an unknown item and which involved establishing a basis from which to put the lexical item to work; and Context Provision strategies, whereby the learner decided upon the linguistic contexts in which the lexical item could be used and how they could fit the item into their existing schemata. The factor that contained the highest number of strategies and onto which the strategies were most heavily weighted was the Cognitive Action factor. This factor comprised the following eight strategies, four of which derived from the top ten most identified with strategies (denoted with an asterisk):

1. Use the item in a sentence as soon as possible
2. Connect the item to a personal experience I have had*
3. Think of situations in which to use the item*
4. Think of words/phrases in English that have a similar/opposite meaning
5. Repeat the item in my head to help me remember it*
6. Remember the word/phrase to ask someone later for the meaning*

7. Think of a picture in my head to help remember the meaning
8. Imagine how the item is spelled

What might explain the prominence of identification with Cognitive Action strategies (cognitive strategies that involve using or preparing to use an newly encountered lexical item) in informal L2 contact? One possible explanation lies in the availability of these eight strategies to a learner regardless of the scenario in which they encounter an unknown word or phrase. That is to say, the solitary nature of these strategies means that they can be applied in both interactive and individual L2 contact scenarios. Moreover, if the Cognitive Action strategies are manifested in interaction, they none of them require the involvement of the interlocutor. Goldstein and Kim (2006) found that the strongest predictors of participation in a study abroad experience were factors related to language socialisation, such as lower ethnocentrism, prejudice and apprehension over intercultural communication, and higher tolerance of ambiguity. This finding denotes that the applicability of the Cognitive Action strategies may not only be broad in terms of context, but also in terms of the learner: it is possible that these strategies are equally particularly attractive to learners who are keen to participate in informal L2 contact and to those who are not.

Another possible reason for the prominence of Cognitive Action strategies (strategies that involve using or preparing to use an newly encountered lexical item) is that those Cognitive Action strategies that also appeared in the top ten most identified with strategies (numbers 2, 3, 5 and 6) are unrelated to proficiency level. The ability to, for example, connect a word to a personal experience or repeat it in your head is independent of L2 proficiency, whereas the success with which a learner is able to, for example, use a new item in a sentence as soon as possible or think of its possible synonyms and antonyms, relies upon their level of linguistic knowledge and/or skill. As such, proficiency-mediated

strategies may be less attractive to lower proficiency learners and therefore less strongly identified with by them.

There was a significant difference in identification with vocabulary-related strategic behaviour between participants undertaking their SAE in Oxford and in London, with participants in Oxford identifying more highly with the strategies from part 2 of the questionnaire than their counterparts in the capital. An independent-samples *t*-test (c.f. 5.5.3) further revealed that there was a significant difference between the higher mean proficiency level of the participants in Oxford than the participants in London. Therefore, the location group that had a higher proficiency level identified more strongly with vocabulary-related strategies, and this finding provides tentative yet suggestive evidence in support of the threshold hypothesis that holds that the lower your proficiency level at the outset of a SAE, the greater your linguistic gain as a result of it.

Another link between proficiency and identification with vocabulary-related strategic behaviour was found: the medium stay participants identified significantly more strongly with the strategic behaviour items than the other two groups. In light of this finding, an ANOVA on proficiency level by length of stay was conducted (c.f. 5.5.4) to find that there was a significant difference between the length of stay groups. Post-hoc comparisons indicated that the mean proficiency level of the medium stay group was significantly higher than that of the long stay group. Here again the study found that the group with the highest proficiency level was also the group that identified most highly with vocabulary-related strategies.

Why might higher proficiency learners identify more strongly with vocabulary-related strategic behaviour in informal L2 contact than their lower proficiency counterparts? A number of other studies have also found a similar link, but rather related to the frequency

of strategy deployment (Ahmed, 1989; Green & Oxford, 1995; Oxford & Ehrman, 1995; Lawson & Hogben, 1996; Park, 1997; Bremner, 1999; Wharton, 2000; Shmais, 2003). It is possible that the more an action is manifested by an individual, the more they identify that action as being 'like them', and this in turn could explain the higher identification with strategies by the groups that comprise the highest proficiency level of learners. In other words, frequency of use may be positively correlated with strength of identification.

6.3 INFORMAL SECOND LANGUAGE CONTACT, VOCABULARY-RELATED STRATEGIC BEHAVIOUR AND VOCABULARY GAIN IN A STUDY ABROAD CONTEXT

Is there a relationship between vocabulary-related strategy use in informal L2 contact and vocabulary gain in a study abroad context?

The findings with reference to informal second language contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context are discussed in four cumulative sections: first, the findings relating solely to the dependent variable are discussed, then the findings relating to informal contact and vocabulary gain, followed by a discussion of the findings pertaining to strategic behaviour and vocabulary gain, and finally is provided a discussion of the findings relating to informal contact, strategic behaviour and vocabulary gain.

6.3.1 VOCABULARY GAIN

The sample as a whole made a statistically significant gain between time 1 and time 2 on receptive, productive and overall vocabulary knowledge as tested by the vocabulary measure in this study. However, the effect size of this gain was small (c.f. 4.1.3.1a). Conversely, in their study of the relationships between communication strategies, L2

contact and automaticity of lexical knowledge, Segalowitz, Freed, Collentine, Lafford, Lazar and Díaz-Campos (2005) found large effect sizes for the gains of their study abroad participants with regard to the speed and efficiency of the sample's L2 visual word recognition. However, as reported in section 5.5.1, in the present study there was a large effect of SAE on the overall vocabulary gain of the long stay group. Segalowitz *et al*.'s (2005) SAE participants stayed in the target language community for thirteen weeks. In the present study the participants in the long stay group were in the UK for 16-20 weeks. Thus, in terms of impact on vocabulary knowledge of SAE, I postulate a *Duration Threshold Hypothesis*: that is to say, it may be the case that up to a certain point in time in the target language community, possibly around the three-month mark, the impact of being in that community on vocabulary knowledge is small, yet beyond that timeframe the impact becomes significantly larger. More evidence in support of a duration threshold with regard to SAE 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. Interestingly, their long stay group stayed in the target language community for as long as or longer than Segalowitz *et al*.'s (2005) SAE participants and the long stay participants in the present study.

Why might there be a duration threshold with regard to gains in vocabulary knowledge in a study abroad context? One potential explanation is strongly allied to the literature discussed in section 2.2.4 on implicit and explicit vocabulary acquisition. If, as posited by Swain's (1985) output hypothesis, production of language 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 study abroad context is less facilitative in the short term of the conditions necessary for

vocabulary knowledge to develop. 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 informal L2 contact scenarios most highly identified with by the sample in the present study (detailed in section 4.2.3) 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 or engaging in short exchanges with cohabitants) and as such these scenarios may not offer the SAE 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 might have developed more confidence to try to manipulate the L2 contact scenarios they encounter in order to facilitate their use of recently encountered lexis. For example, one of the OWLS interview participants reported that during his SAE he bought a packet of cigarettes regularly from the newsagent's between his language school and the closest London Underground station to the school. He told me how, over time, his interactions with the shopkeeper had gone from simply asking for the cigarettes and being told how much to pay, to a cultural exchange of sorts, whereby the two men discussed their respective countries (Spain and India) and struck up what sounded like a genuine friendship. Thus, from a highly limited functional interactional exchange, time had facilitated access to a series of scenarios in which the participant was able to use a variety of lexical items on a wide rage of topics. It may also be the case that as a SAE learner's personal relationships with L2 speakers (e.g. their host family, their fellow students) become deeper and more familiar over time, so they become more comfortable with producing language with those interlocutors and testing their fledgling lexical hypotheses.

The long stay group outperformed the medium stay group on overall and receptive vocabulary gain. These findings support 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, and as here discussed, Ife *et al.* (2000) found what they described as tentative evidence in support of the positive impact of greater length of stay on vocabulary development. DeKeyser (2014) argues for the interplay between length of stay and proficiency level; in other words, 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. As reported in 5.5.4, the medium stay participants had a significantly higher mean proficiency level 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.

6.3.2 INFORMAL CONTACT AND VOCABULARY GAIN

The correlation analyses of informal L2 contact and vocabulary gain found no statistically significant relationship between the two. The findings of the regression analyses on informal contact and vocabulary gain (c.f. 4.3.2) indicate that the model as a whole did not significantly predict vocabulary gain. Only one predictor variable in the model, length of stay, made a statistically significant contribution: belonging to the medium stay group predicted lower receptive and overall vocabulary gain but had no significant relationship to productive vocabulary gain. As revealed in 5.5.4, the medium stay group had a significantly higher mean proficiency level than the long stay group. Therefore, once more the 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 (11-15 weeks).

What might explain the lack of impact of informal L2 contact on vocabulary gain in this study? It may be the case that the extent to which the participants identified with the informal L2 contact scenarios was insufficient to reveal a statistically significant relationship between these variables. That is to say, the sample may not have had enough informal 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 size. As mentioned in section 6.3.1, the informal L2 contact scenarios most highly identified with by the sample involved either receptive activity or interaction which was likely to be limited in terms of the lexis it involves and as such these scenarios may not have offered the participants many opportunities to use recently encountered lexical items.

Another hypothesis is that the informal L2 contact the participants experienced in this study had an impact on other dimensions of vocabulary knowledge, such as depth and/or automaticity, rather than on breadth. Certainly, Fitzpatrick (2012) found that study abroad had a complex and non-linear impact on her participant's acquisition of individual lexical items, and study abroad 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, if temporal, financial and manpower resources allow.

6.3.3 STRATEGIC BEHAVIOUR AND VOCABULARY GAIN

There existed a significant but small, negative relationship between strategic behaviour and receptive and overall vocabulary gain, with higher levels of identification with strategic behaviour associated with lower receptive and overall vocabulary gain. However,

the correlation analysis that revealed this finding took the unfactored strategic behaviour variable as the independent variable and therefore treated the strategies as if they were conceptually homogenous. As this approach counters the theoretical standpoint that strategies are disparate and largely context-specific and most effectively operate in combinations, this finding was discounted in favour of the findings of the correlation analyses conducted using factored strategic behaviour.

In terms of the factored strategic behaviour data, there was a small, positive relationship between identification with strategies that loaded onto the Cognitive Action factor and receptive and overall vocabulary gain, indicating that higher identification with Cognitive Action strategies was linked to higher receptive and overall gain. This finding is encouraging given that the Cognitive Action factor comprised the highest number of and most heavily loaded strategies. Moreover, that the Cognitive Action factor comprised the highest number of strategies and is positively correlated with gain is supportive of Schmitt's (1997) assertion that successful language learners utilise a variety of vocabulary-related strategies.

As detailed in 6.2.2, the eight strategies in the Cognitive Action factor comprised Individual Determination, Memorisation, and Metacognitive strategies. They all presupposed that the learner has decided to try to *use* the lexical item in question. The majority of the strategies involved preparation for using the item and all were solitary and cognitive. This finding mirrors those of a number of other scholars: Gu and Johnson found that a group of strategies with similar characteristics correlated positively with vocabulary scores: their 'activation of newly learnt words' (1996:643) category, which comprised actions taken by the learner that involved *using* a newly-learned word, such as using the lexical tem in an imagined or real situation, or looking for examples of or opportunities to use the item. Similarly to the present study's Cognitive Action strategies, Gu and

Johnson's (1996) five Activation strategies were all solitary and predominantly cognitive. Kojic-Sabo and Lightbown (1999) also found that strategies of Independence, which they defined as effort on the part of the learner to encounter and *use* new words/phrases, in tandem with a preference for learning vocabulary in informal L2 contact, were associated with high levels of vocabulary learning success. In another similar finding, Fan (2003) revealed that strategies involving reviewing and *using* newly encountered lexical items were among the most highly and positively correlated with vocabulary learning. All of these findings considered together provide convincing evidence of the value to vocabulary gain of vocabulary-related strategies that involve using or preparing to use a newly encountered lexical item. One could argue, therefore, that these findings combined with the findings of the present study provide robust support for Swain's (1985) output hypothesis, in that a powerful 'trigger' in the acquisition of vocabulary seems to be the conscious use of it.

Correlation analysis revealed a relationship between identification with the Establishment strategies and productive vocabulary gain: higher identification with strategies that loaded onto the Establishment factor was correlated with lower productive vocabulary gain. As discussed in 5.2.3.2, the Establishment factor was comprised of six strategies, all of which involved establishing a basis from which to put the item to work. The strategies were, in order of heaviest to lightest weighting:

1. Write down the item and its meaning immediately
2. Draw a picture/diagram to help me the remember the item and its meaning
3. Set myself goals for learning/using the item
4. Ask the speaker to translate the item into my native language
5. Say the item aloud to help me remember it
6. Use a bilingual dictionary to find the translation of the item in my language

Although the strategies in this factor comprised a mix of Determination, Memorisation and Metacognitive strategies, they none of them involved use of the item in context in the L2. It is possible that this mismatch between the type of knowledge application described in these strategies and the type of knowledge tested in part 2 of the vocabulary test accounted for the negative correlation between the two. In other words, all of these strategies would most likely be deployed soon after first contact with an unknown lexical item, and therefore participants who were identifying strongly with them would most likely be at the stage of acquiring receptive as opposed to productive knowledge of the item.

Conversely to the negative correlation between Establishment strategies and productive gain in the present study, Fan (2000) found a positive correlation between productive vocabulary scores and seven similar strategies. Unfortunately, although Fan (2000) stated that her vocabulary strategies questionnaire measured both frequency of use and perceived usefulness, we are not told which of these variables was correlated with the active vocabulary scores. The seven strategies were:

1. I learn new words from coursebooks, handouts, or anything written in English from inside school
2. I increase my English vocabulary by reading stories, newspapers, magazine etc. outside of class
3. I play games in English to learn more new words
4. When I meet new words in reading, I guess their meaning and then look up the dictionary
5. I group words that are related to help myself remember them
6. I group together words and expressions that are used in a certain situation e.g. Bank: cash, cheque, open an account etc.
7. I repeatedly say the word in my mind

There are a number of possible reasons for the divergent findings with regards productive vocabulary between the present study and that of Fan (2000). Firstly, the vocabulary strategy questionnaires administered in each were measuring different constructs (identification with vs. frequency of use/perceived usefulness). Secondly, Fan's (2000) strategies were not discrete (for example, strategy four clearly describes two strategies deployed as a chain), are not item, context or task specific (e.g. all bar the final strategy listed relate to vocabulary in general as opposed a specific word or phrase encountered), and do not make clear the relationship between action and intention (e.g. what information is the learner hoping to find in the dictionary?). Most of the seven strategies that Fan (2000) found negatively to correlate with active vocabulary pertain to first contact with a lexical item (similarly to those strategies negatively correlated to productive gain in the present study): as such, a positive relationship between the seven strategies and the development of productive ability, which is widely believed to develop after receptive ability (Pignot-Shahov, 2012), is questionable.

The findings of the regression analyses which took strategic behaviour, length of stay and location as the predictor variables indicated that for receptive vocabulary gain, three of the strategic behaviour factors were significant contributors: Appraisal strategies predicted positive receptive gain whereas Visual Reference and Establishment strategies predicted negative gain. Additionally, the model found a significant contribution of length of stay, with belonging to the medium stay group a predictor of lower receptive gain. The Appraisal factor comprised six strategies in which the learner is making judgments about the lexical item in terms of what it means and whether/how they should learn it.

1. Try to guess the meaning
2. Ask the speaker to give me a synonym of the item in English
3. Decide whether or not I want to remember/learn the item

4. Work out what part of speech the item is (e.g. noun/verb/adjective)
5. Ask the speaker to explain (in English) the meaning of the new item
6. Decide the best way to learn the word/phrase

It is perhaps unsurprising that the majority of the strategies in the Appraisal factor are independent of context, and all are available to a learner in the kind of L2 contact most identified with most highly by this sample (Interactive scenarios).

For productive vocabulary gain, neither the initial nor the factored model was able to reveal any explanatory power. However, in a finding which mirrored that of the correlation analyses, the factored regression model revealed a significant contribution of Establishment strategies, whereby stronger identification with these strategies predicted lower levels of productive gain. For overall vocabulary gain, when strategic behaviour was represented in the model by the factored strategic behaviour variables, identification with Visual Reference strategies and belonging to the medium stay group both predicted lower overall gain. All of the strategies in the Visual Reference factor describe a solitary endeavour that makes use of visual information about the item in order better to determine its meaning and use:

1. Use an English-only dictionary to find the item used in a sentence
2. Use an English-only dictionary to find the meaning of the item
3. Use any available pictures/body language to help me understand the item

In terms of the first two strategies in this factor, one could argue that these behaviours are limited for the most part to informal L2 contact to activities which involve reading, yet in the present study the most highly identified with informal contact scenarios were Interactive. Similarly, the strategies in the Establishment factor are also highly context dependent and perhaps less available to a learner who is engaged in interaction. Fan found

that the lowest proficiency learners in her study reported more frequent use of the strategy "I use the dictionary to find out the appropriate usage of the word" (2003:235). Conversely, she also found that the most proficient learners reported more frequent use of "using the dictionary to find out the context meaning of the new word" (2003:235). Unfortunately, there is significant overlap between these two strategies in that the 'context meaning' of a word in a dictionary is often an example of its appropriate usage. Dictionary-based strategies are perhaps the most heavily context-dependent of all of the vocabulary-related strategies because their use is very often prohibited by the confines of physical environment or the strictures of interactional norms. As such, to measure their relationship to generalised outcomes may indeed lead to somewhat confusing findings.

It may be the case that there was some qualitative difference that accounted for the finding that belonging to the medium stay group predicted lower receptive and overall vocabulary gain. The short stay group may have felt a sense of urgency (similar to the hypothesised reason behind their greater identification with informal contact scenarios) which underpinned their gain, and the long stay group may simply have had more exposure and more frequent subsequent exposures to words and phrases. However, this finding may also once more be attributable to the higher proficiency of the medium stay group in comparison with the long stay group.

6.3.4 INFORMAL CONTACT, STRATEGIC BEHAVIOUR AND VOCABULARY GAIN

Stronger identification with informal L2 contact scenarios was found to have a medium-strength, positive correlation with stronger identification with vocabulary-related strategic behaviour. The most likely reason for this finding is that learners who have more contact

with the L2 in informal settings have more opportunity to deploy those vocabulary-related strategies that are available to them in informal settings.

The findings of the regression analyses in the present study which took informal contact, strategic behaviour, length of stay and location as the predictor variables indicated that for receptive vocabulary gain, stronger identification with Cognitive Action and Appraisal strategies and Narrative scenarios predicted higher gain. The Narrative factor comprised six scenarios all related to the production or comprehension of narratives in some way: they each involved either reception of narratives (through watching films and using subtitles to enhance comprehension of the narrative) or represented situations in which a learner might narrate their experiences to others:

1. I write emails in English outside of class
2. I read subtitles while I am watching TV or movies in English outside of class
3. I listen to English movies and videos outside of class
4. I write personal notes or letters in English outside of class
5. I speak my native language to non-native speakers of my native language
6. I speak English to native or fluent speakers of my native language

There is an interesting relationship between Cognitive Action strategies, Appraisal strategies and Narrative scenarios in that all of the Narrative scenarios (excepting perhaps scenario 5) allow for the vast majority of the Cognitive Action and Appraisal strategies to be deployed. That is, the Narrative scenarios facilitate a larger number of strategies than the other informal contact factors, and for a number of different reasons. Firstly, the Narrative scenarios are largely independent of the strictures of interactional norms, and those scenarios that do involve interaction carry a lower pressure level because the interaction is either (1) not the main focus of activity (i.e. if watching TV/movies), (2)

being carried out in the L1, or (3) carried out with the 'safety net' of being able to codeswitch to the L1 to repair any breakdown in communication. Conversely, the Interactive scenarios restrict many of the vocabulary-related strategies available because to deploy them would be to disrupt the interactional flow. The Literate scenarios are more similar to the Narrative scenarios in that they comprise a range of interactive and solitary endeavours; however, one could argue that the Narrative factor comprises scenarios that are more personalised to the learner. That is to say, the Narrative scenarios pertain more strongly to enjoyment and relationship building than the Literate scenarios, which in turn appear to be less personal and tend more to the obtainment and management of information. From a sociocognitive perspective, Atkinson (2010) states that language learners pay most attention to that which really matters to them; to language learning opportunities that offer the greatest potential in terms of their social integration and survival. From this argument can be extrapolated that a more personalised context begets deeper processing of language that the learner has decided to attend to. As such, I contend that it is the combination of the these two key qualities inherent in the Narrative scenarios, the fact that they facilitate the deployment of a greater range of strategies and are more highly personalised to the learner, that accounts for their positive relationship to vocabulary gain.

Once more, stronger identification with Visual Reference and Establishment strategies and Interactive scenarios, and belonging to the medium stay group were found to predict lower gain. The model had no explanatory power over productive vocabulary gain. In terms of overall vocabulary gain, identification with Cognitive Action and Visual Reference strategies and Narrative scenarios had a positive relationship with gain, whereas identification with Establishment strategies and belonging to the medium stay group once more predicted lower levels of gain. Similarly to the Cognitive Action strategies, all of the

Visual Reference strategies are facilitated by one or more of the Narrative scenarios, and when deployed in Narrative scenarios, either as individual strategies or in combinations including Cognitive Action strategies, the Visual Reference strategies serve to enhance the socio-affective standing of the learner on a SAE.

6.4 CONTRIBUTION TO THEORY AND PRACTICE

In this section is discussed the unique contribution of the present study to SLA theory and research practice.

6.4.1 CONTRIBUTION TO THEORY

Before this study (and as discussed in section 2.3) strategic behaviour had been conceptualised from a cognitive and then, more recently, from a sociocultural theoretical perspective. Over the past decade many researchers (e.g. Macaro, 2004; Grenfell & Harris, 2013), regardless of epistemological stance, have argued that strategic behaviour is context dependent and as such that strategic behaviour cannot fully be understood separate from the context in which it is manifested. In simple terms, strategies have been theoretically bound together with context.

The unique contribution of the present study to the theoretical conceptualisation of strategic behaviour is that the data yielded by the OWLS tool has demonstrated the pivotal role played by *intention*; that is to say, the relationship between vocabulary-related strategic behaviour and context is wholly dependent upon the intention of the learner in that context. Without the intention to comprehend or acquire a newly encountered lexical item, regardless of the potential of a given scenario to facilitate varied and numerous strategic responses, there is no relationship between context and vocabulary-related

strategic behaviour. *Intention* is what makes the difference, and this innovation is the original contribution of this study to SLA theory.

6.4.2 CONTRIBUTION TO PRACTICE

The 'toolbox' for researching vocabulary-related strategic behaviour comprises direct and indirect methods (c.f. section 2.3.4.2). As stated by McDonough (1995), direct methods pertain to participants providing a self-report. Direct methods can be concurrent to a task (e.g. think aloud protocol) or retrospective (e.g. stimulated recall). Indirect methods involve participants stating the extent of their identification/agreement with or the frequency of their experience of researcher-worded statements. The OWLS, therefore, is an innovative addition to the direct methods in the toolbox, whereas part 2 of the questionnaire used in the present study is a novel contribution to indirect methods.

There are a number of more specific unique contributions of this study to strategy-based research practice. Firstly, before the present study no research instruments or analytic tools existed that were focused specifically on strategic behaviour that occurs outside of a formal language learning setting. Three such tools are now in existence: (1) the OWLS instrument; (2) part 2 of the questionnaire (which measured identification with vocabulary-related strategic behaviour that occurs outside of the classroom); and (3) the vocabulary-related strategies taxonomy (from which the items in part 2 of the questionnaire derived). Secondly, the OWLS has resolved the methodological shortcomings of the existing direct research methods in the toolbox: As discussed in detail in Briggs (2014), stimulated recall is affected by the propinquity of the task to the recall of it; in other words, the fallibility of memory may impact the reliability of the data yielded by stimulated recall, and more so the greater the amount of time between task and recall. The OWLS overcomes this shortcoming by employing *simulation* rather than stimulation so that the participant does

not recall their thought processes but rather reports on them concurrently, similarly to think aloud protocol. However, one criticism of think aloud protocol put forward by White, Schramm and Chamot (2007) is that simultaneous verbalisation of thought and attending to a task is a highly complex endeavour and as such a participant's ability to report their thought processes accurately/in detail and/or to complete the task at hand may be limited. The OWLS overcomes this shortcoming because it does not involve a task: participants are not required by the OWLS to achieve a goal or reach a conclusion of any kind, but simply to recognise each scenario and verbalise their thought processes in response.

Another unique contribution made by the OWLS instrument to vocabulary-related strategy-based research practice is that it has the potential to reveal the relative influence of task, context, intention, L1, interlocutors, input mode and L2 lexis on strategic behaviour (the relationships between which are illustrated in figure 6.1). That is to say, researchers can modify the OWLS to determine whether and how a change in one of these independent variables yields a change in the strategic responses to it. Moreover, the OWLS is ripe for adaptation to other language learning settings, such as the classroom, the workplace, and online.

To summarise, the present study has made a unique contribution to SLA practice with regard to vocabulary-related strategic behaviour in four ways: The present study has (1) contributed to both direct and indirect methods of investigating vocabulary-related strategic behaviour; (2) yielded the only research tools in existence which were designed specifically for the investigation of vocabulary-related strategic behaviour in informal contexts; (3) produced a direct research tool (the OWLS) which has overcome the methodological shortcomings of the direct tools available to date; and (4) created a

research tool (the OWLS) capable of revealing the relative impact of a range of variables on vocabulary-related strategic behaviour.

CHAPTER VII

CONCLUSIONS

The present study is founded on the body of literature on informal second language contact, vocabulary-related strategic behaviour, and vocabulary gain in a study abroad context. As argued in Chapter II, three main gaps in knowledge arise from analysis of the literature: (1) the evidence of informal L2 contact is largely unreliable, ungeneralisable, or both; (2) the evidence of vocabulary-related strategic behaviour in informal L2 contact is neither context nor task specific; and (3) there is no present evidence of the relationships between informal L2 contact, vocabulary-related strategic behaviour and vocabulary again in a study abroad context. The present study has built upon these limitations and in doing so has made an original contribution to the body of knowledge of SLA research. In this chapter is provided a synthesis of these key findings, of the limitations of the study and of its pedagogical implications. Finally, some recommendations for future research are made, and the concluding remarks set forth.

7.1 KEY FINDINGS

7.1.1 INFORMAL L2 CONTACT

The synergy between the findings of this study and those of Briggs (2014), which is the only other study to have taken the same methodological and analytical approach to researching informal L2 contact, provides strong evidence about the kinds of informal L2 contact that study abroad learners in England are experiencing. Firstly, there is a distinction in the minds of study abroad learners between informal contact that is interactive and contact that is solitary. Secondly, informal L2 contact that is interactive is most likely to occur with interlocutors with whom the learner lives or studies. Finally,

study abroad learners most identify with informal L2 contact scenarios that involve them seeking information from external sources, such as an interlocutor, a poster, a website and so on.

7.1.2 VOCABULARY-RELATED STRATEGIC BEHAVIOUR

With reference to vocabulary-related strategic behaviour in a study abroad context this study revealed that, as also found in Briggs (2014), strategy use in informal contact is heavily learner and context dependent. Uniquely to this study, interplay between the learner's L1 and the cultural context of the SAE was found to yield assumptions on the part of the participants about the informal L2 contact scenarios they encountered. Strategies that involve using or preparing to use a newly encountered lexical item were the most highly identified with by the sample. Furthermore, and similarly to the findings of Gu and Johnson (1996), Kojic-Sabo and Lightbown (1999) and Fan (2003), strategies of use and preparation for use were associated with higher levels of vocabulary gain.

7.1.3 INFORMAL CONTACT, VOCABULARY-RELATED STRATEGIC BEHAVIOUR AND VOCABULARY GAIN IN A STUDY ABROAD CONTEXT

Identification with scenarios that allowed for the greatest range of vocabulary-related strategies to be deployed, and that were personalised to the learner was predictive of higher vocabulary gain. In other words, learners who gained the most vocabulary knowledge were those whose informal contact with English was meaningful and did not restrict their strategy use. Participants who gained less vocabulary knowledge identified more highly with L2 contact scenarios that restricted the number and/or range of vocabulary-related strategies that they could employ. Similarly, higher identification with strategies that are less context-dependent (i.e. strategies that can be used in any context) was predictive of higher vocabulary gain, and identification with strategies that were more

context-dependent was predictive of lower vocabulary gain. Participants in the medium stay group, who had a significantly higher proficiency level than the long stay participants, gained less vocabulary knowledge during their SAE.

7.2 LIMITATIONS

A number of limitations are hereby set forth that should be considered with reference to the key findings and main arguments of this study.

7.2.1 SAMPLING

The study employed purposive non-probability sampling. As such it cannot reliably be argued that the findings here reported are representative of the population; nor can the findings statistically be generalised to the population. Moreover, it was not possible in this study to control for the effects of the instruction that the sample was receiving as part of their study abroad programme. The heterogeneity of the sample may have limited the statistical power of the quantitative analyses conducted: particularly with reference to identification with informal L2 contact it may be the case that the participants' behaviour was so divergent from one another that patterns in the data failed to be revealed.

7.2.2 MEASURES

The test used to measure the dependent variable of vocabulary gain in this study represented the most logical choice given the available options and with reference to the sample and the testing load placed upon them. Ideally, a simple L2-L1 vocabulary test would have been employed, whereby the levels test definitions occurred in participants' L1 or the L1 equivalent of the target words could be selected; this would have meant that the responses would not be reliant on comprehension of the possible answers. However, due to the wide variety of L1s sampled, a L2-L2 test was necessary.

As a measure of vocabulary breadth, Nation's (1990) levels test may not be entirely accurate. The frequency levels are derived from and validated against native-speaker corpora, and the major theoretical assumption underlying the test is that the most frequent vocabulary words are acquired first. However, as a result of non-natural input and exposures to English, L2 learners may remember a large quantity of low frequency words but lack the basic knowledge of high frequency words. However, as the test was used as a measure of gain as opposed to vocabulary size, this limitation was neutralised. Laufer and Nation's (1999) productive vocabulary levels test is not entirely a measure of productive knowledge because in giving the first letters of the target word, receptive knowledge is employed to recognise that word-part. However, a true vocabulary size test of productive knowledge is so difficult to design that as yet one does not exist (Gu, 2014: personal correspondence).

7.2.3 PROFICIENCY LEVEL

It may have been the case that some participants' ability to describe their strategic behaviour in the OWLS interviews was inhibited by their proficiency level in English. Indeed, some of the OWLS interview transcript extracts suggest that at times some participants were unable to convey their understanding of the simulated scenario or comprehension of a target word. However, in conducting the interviews much use was made of body language, gesture and physical reference to the computer screen on which the OWLS was played; so much so that, once a participant had grasped the format of the interview, no insurmountable breakdowns in communication occurred.

7.2.4 TIME AND RESOURCES

As a result of funding strictures, the planning, execution and reporting of this study was constrained to within a time period of three years. As such, compromises were made: the

use of the OWLS instrument in semi-structured interviews was limited to a relatively small subsample and it was not possible to interview the sample about their responses to the informal L2 contact and strategic behaviour items in the questionnaire.

It should be noted that this study represents the very first occasion on which the relationships between informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context have been investigated, and as such the study does not provide the definitive solution to researching these variables but rather opens the door for future research to follow.

7.3 PEDAGOGICAL IMPLICATIONS

Private language institutions sell their study abroad programmes on the promise of informal L2 contact. Entering the term 'study abroad UK' into a search engine invariably results in a bombardment of emotive language such as 'The world is waiting' and 'Invest in your dreams'. If a study abroad learner elects to live with a UK host family during their SAE, they are in many cases promised breakfast and dinner with the host family during the week, and all meals with them during the weekends. Unfortunately, in the course of data collection for this study it was not uncommon to encounter students who had little to no interaction with their host family: in one extreme case the host family fridge was padlocked to prevent the student from using it. Promises are also made to potential students from the outset about the impact of L2 contact on their linguistic development, e.g. 'pick up expressions with greater ease' and 'return home with a more authentic accent'. However, for the most part, the formal language instruction that SAE learners receive is no different from what is delivered in EFL contexts. Certainly, little differentiation is made in initial or advanced ELT teacher training between what should be taught in EFL and ESL settings. I argue, therefore, that study abroad language institution

curricula should prepare learners for the informal L2 contact they are likely to experience during the SAE, ensure that learners are provided with the informal contact they are promised, raise their awareness of the strategies that will be available to them in the informal L2 contact that they are likely to encounter, and provide practice in manipulating the most likely scenarios with the available vocabulary-related strategies. One could argue that to continue to sell courses on the promise of informal contact yet fail to provide that contact or to give instruction on how to manipulate it for linguistic gain is tantamount to exploitation.

The findings of this study suggest that strategy instruction should be embedded in context; that is to say, that where strategy instruction takes place the learner is invited and guided to link the strategy or combination of strategies to the situations and tasks which facilitate it. This implication holds true not only in the study abroad context but in all language learning settings. Moreover, this study has shown that the interplay between a learner's L1 and the culture of the target language community has an impact on the vocabulary-related strategies they choose from, suggesting that strategy instruction for a SAE cannot be generalised across L1s and across SAE contexts: it must be bespoke to the learner and to the SAE context they are due to meet. This conclusion raises an important question: How might private language institutions go about preparing study abroad learners for strategic behaviour during the SAE in light of group differences such as L1? Most study abroad classes are multilingual and as such, a L1-based bespoke instructional approach after arrival to the UK is rarely possible. However, classes on 'British culture' are often taught on study abroad programmes and this topic might usefully be exploited to explore different cultural expectations of informal contact with English and how those expectations might influence learning behaviour. Another possible approach involves the online engagement of the learner in considering informal L2 contact before they have left their home country.

It is my experience that UK private language institutions have not yet embraced Virtual Learning Environments to the same extent as higher education institutions: platforms such Moodle or Blackboard might usefully serve to deliver L1/nationality/proficiency-specific input with regards the impending sojourn abroad.

Another pedagogical implication of this study related to language learning strategy pedagogy is that strategies that involve the *use* of the newly encountered word seem to be most beneficial to learners in informal L2 contact in terms of maximising that contact for vocabulary gain. This is because the findings of this study and others (c.f. 6.3.3), considered together, provide convincing evidence of the value to vocabulary gain of vocabulary-related strategies that involve using or preparing to use a newly encountered lexical item in informal L2 contact. English language teachers in all contexts should be apprised of this finding in order to help their learners to maximise whatever type or amount of informal L2 contact they experience. As stated almost a decade ago by Gu and Johnson (1996:659), "Vocabulary learning should hence aim toward vocabulary in action". Here the scope of the OWLS instrument could usefully be extended beyond that of a research tool; without a great deal of modification it could serve as a pedagogic aid to help train language learners to make effective strategic choices in a variety of informal scenarios.

7.4 RECOMMENDATIONS FOR FUTURE RESEARCH

A number of recommendations for future research can be made as a result of the study here reported.

7.4.1 STUDY ABROAD EXPERIENCE

The current evidence on informal L2 contact on a SAE is mostly derived from learners who are engaged in an academic SAE at a tertiary level: hardly surprising given that the majority of academic researchers work at universities. However, this means that the population of SAE learners who study at private language institutions are under-represented in study abroad research. I call for this population to receive more attention in future: the cost of one week's study at a private language school can be equal to the cost of an academic term at a UK university, and many students leave the target language community with only a certificate of competence or course completion issued by their language school. Moreover, the British Council Accreditation UK scheme, the quality assurance scheme for ELT provision, is voluntary and conducts school inspections only once every four years. As such, students who study abroad for language acquisition are in greater danger of exploitation from the institutions they study with than are their academic study abroad counterparts. Study abroad for language acquisition is under-regulated, but it need not be under-researched.

7.4.2 INFORMAL L2 CONTACT

DeKeyser (2014:317) stated that future research into study abroad might usefully document how informal opportunities for L2 acquisition present themselves and describe in fine-grained detail how learners take advantage of them in light of the interaction between the learner and scenario. This study has heeded his call. The use in future research of a more homogenous sample than was employed in the present study may have the potential to reveal qualitative differences between learners that account for any differences in their identification with certain types of informal L2 contact scenarios. Future research

on informal L2 contact in a study abroad context might usefully focus on either interactive or individual scenarios, or constitute a comparison of the two.

It would be interesting and useful for future research to define the informal L2 contact scenarios experienced by learners of English as a foreign language in their home countries in order to extend the scope of the present study to learners outside of the target language community and define the kinds of informal contact which each respective learning context has the most and least potential to provide.

7.4.3 VOCABULARY-RELATED STRATEGIC BEHAVIOUR

The sheer weight of evidence found in the present study and in Briggs (2014) in support of the context-specific nature of strategic behaviour calls for an end in SLA research to the cumulative treatment of data pertaining to strategy use. If the data itself has not revealed any underlying links between strategies (i.e. as yielded by factor analysis in the present study), then there seems little value in making generalisations that are themselves based on generalisations.

It would be useful to have a taxonomy of vocabulary-related strategies in which the strategies are ranked or grouped together based on their dependence upon or independence from learner or contextual features. If such a categorisation existed, researchers could begin to investigate the dependency between different L2 contact scenarios and the strategies deployed by learners within them. I would argue that a taxonomy of this nature would only be truly useful if it were bespoke to the very population a study is seeking to investigate: that is, the ranking or grouping of the strategies by context-dependency would most usefully be conducted by the study abroad learners themselves.

7.4.4 VOCABULARY GAIN

As discussed in 6.3.2, besides receptive and productive growth there are other aspects of vocabulary gain, such as depth and automaticity of knowledge, which may be impacted as a result of SAE. In fact, it may be the case that natural and more informal exposures to English during SAE result in more depth and automaticity than breadth. As such, future research into the impact of study abroad on vocabulary knowledge might usefully include pre and post measures of breadth and automaticity in tandem with testing for gains in size.

7.5 CONCLUDING REMARKS

This study has yielded the following main arguments:

1. Study abroad learners in England identify most highly with informal L2 contact scenarios that involve them seeking information from external sources (e.g. people, websites, timetables).

2. The vocabulary-related strategies most highly identified with by study abroad learners in England pertain to the *use* of a newly encountered lexical item.

3. Vocabulary-related strategic behaviour in informal L2 contact is too heavily dependent upon learner and contextual features to be considered cumulatively, and as such strategic behaviour in informal settings should be investigated either as a purely qualitative endeavour, or based on data-derived subdivisions, or as a combination of the two.

4. There is a relationship between informal second language contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context: informal

contact scenarios that are less strategically prohibitive and vocabulary-related strategies that are less context-dependent are positively related to vocabulary gain.

These arguments comprise the original contribution of this thesis beyond that which was previously known. This study has addressed the gap in knowledge about the relationships between informal L2 contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context.

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 (2), 352-364.

Adams, R. (2006). Language learning strategies in the study abroad context. In M.A. Dufon & E. Churchill (Eds.), *Language learners in study abroad contexts.* Clevedon, England: Multilingual Matters (pp. 259–92).

Adolphs, S. & Schmitt, N. (2003). Lexical coverage and spoken discourse. *Applied Linguistics*, 24 (4), 425-38.

Ahmed, M.O. (1989). Vocabulary learning strategies. In P. Meara (Ed.), *Beyond words.* London: British Association for Applied Linguistics, in association with Centre for Information on Language Teaching and Research (pp. 3-14).

Aizawa, K. (2006). Rethinking frequency markers for English-Japanese dictionaries. In M. Murata, K. Minamide, Y. Tono & S. Ishikawa (Eds.), *English lexicography in Japan.* Tokyo: Taishukan Publishing Company (pp. 108-119).

Albrechtsen, D., Haastrup, K. & Henriksen, B. (2008). *Vocabulary and writing in a first and second language: Processes and development*. Basingstoke: Palgrave Macmillan.

Allen, H.W & Herron, C. (2003). A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals*, 36 (3), 370-385.

Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations. In A.P. Cowie (Ed.), *Phraseology: theory, analysis and applications.* Oxford: Oxford University Press (pp. 101-122).

Altenberg, B. (1990). Speech as linear composition. In G. Caie, K. Haastrup, A.L. Jakobsen, J.E. Neilsen, J. Sevaldsen, H. Spect & A. Zettersten (Eds.), *Proceedings from the Fourth Nordic Conference for English Studies, Vol. 1*. Copenhagen, Demark: Department of English, University of Copenhagen (pp. 133-143).

Amuzie, G.L. & Winke, P. (2009). Changes in language learning beliefs as a result of study abroad. *System*, 37 (3), 366-379.

Anderson, N.J. (2008). Metacognition and good language learners. In. C. Griffiths (Ed.), *Lessons from good language learners.* Cambridge: Cambridge University Press.

Aston, G. & Burnard, L. (1998). The BNC handbook: exploring the British National Corpus with SARA. Edinburgh: Edinburgh University Press.

Atkinson, D. (2010). Extended, Embodied Cognition and Second Language Acquisition. *Applied Linguistics,* 31, 599-622.

Avello, P., Mora, J.C. & Pérez-Vidal, C. (2012). Perception of FA by Non-native Listeners in a Study Abroad Context. *Research in Language*, 10 (1), 63-78.

British Association for Applied Linguistics (BAAL) (2000). *Recommendations for good practice in Applied Linguistics student projects.* Available from: www.baal.org.uk/dox/goodpractice_full.pdf [Last accessed 10 January 2012].

Bacon S.M. (1992). The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign language listening. *The Modern Language Journal*, 76, 160-174.

Bacon, S.M.C. & Finnemann, M.D. (1992). Sex differences in self-reported beliefs about language learning and authentic oral and written input. *Language Learning,* 42 (4), 471-495.

Baddeley, A.D. (1997). *Human memory: Theory and Practice* (Revised Edition). Hove: Psychology Press.
Barcroft, J. (2009). Strategies and performance in intentional L2 vocabulary learning. *Language Awareness*,18, 74-89.
Bardovi-Harlig, K. (2002). A new starting point? Investigating formulaic use and input. *Studies in Second Language Acquisition*, 24, 189-198.
Barfield, A. (2009). Following individual L2 collocation development over time. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations.* New York: Palgrave Macmillan (pp. 208-223).
Barfield, A. & Gyllstad, H. (Eds.). (2009). *Researching collocations in another language: Multiple interpretations.* Basingstoke: Palgrave Macmillan.
Baró, A.L. & Serrano, R.S. (2011). Length of stay and study abroad: Language gains in two versus three months abroad. *Revista Española de Linguística 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 (1), 2-3.
Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27 (1), 101-18.
Beglar, D. & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing*, 16, 131-162.
Bell English (2014). *Intensive English* [Webpage]. Available at http://www.bellenglish.com/Book_Now/73537/Intensive-English/ [Last accessed 18 September 2014].
Berry, W.D. (1993). *Understanding regression assumptions.* Beverly Hills, CA: Sage.
Bialystok, E. (1983). Some factors in the selection and implementation of communication strategies. In C. Færch & G. Kasper (Eds.), *Strategies in interlanguage communication.* Harlow, England: Longman, (pp. 100-118).
Biber, D. (1988). *Variation across speech and writing.* Cambridge, England: Cambridge University Press.
Biber, D., Gray, B. & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45 (1), 5-35.
Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English.* Edinburgh: Pearson Education Limited.
Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56, 9-49.
Birdsong, D. & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, 44, 235-249.
Block, D. (2003). *The Social Turn in Second Language Acquisition*. Edinburgh: Edinburgh University Press.
Bogaards, P. (2001). The use of grammatical information in learner's dictionaries. *International Journal of Lexicography*, 14 (2), 97-121.
Bonk, W.J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, 14 (1), 14-31.
Brecht, R., Davidson, D. & Ginsberg, R. (1995). Predictors of foreign language gain during study abroad. In B.F. Freed (Ed.), *Second language acquisition in a study abroad context.* Amsterdam, The Netherlands: John Benjamins (pp. 37-66).

Brecht, R. & Robinson, J. (1995). On the value of formal instruction in study abroad: Student reactions in context. In B.F. Freed (Ed.), *Second language acquisition in a study abroad context.* Amsterdam: Benjamins (pp. 317-334).
Bredo, E. (1994). Reconstructing Educational Psychology: Situated Cognition and Deweyan Pragmatism. *Educational Psychologist*, 29 (1), 23-35.
Bremner, S. (1998). Language learning strategies and language proficiency: Investigating the relationship in Hong Kong. *Asian Pacific Journal of Language in Education*, 1 (2), 490-514.
Bremner, S. (1999). Language learning strategies and language proficiency: Investigating the relationship in Hong Kong. *Canadian Modern Language Review*, 55(4), 490-514.
Briggs, J.G. (2010). *A whole new world of words: Out-of-class contact, strategic behaviour and vocabulary gain in a study abroad context.* Unpublished Masters dissertation. University of Oxford, UK.
Briggs, J.G. (2014). A context-specific research tool to probe the out-of-class vocabulary-related strategies of study abroad learners. *International Journal of Applied Linguistics* [In press]. DOI: 10.1111/ijal.12070.
British Council (2014). *Frequently asked questions: the English language.* Available at http://www.britishcouncil.org/learning-faq-the-english-language.htm [Last accessed 12 August 2014].
Brown, R., Waring, R. & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20 (2), 136-163.
Bryman, A. (2004). *Social Research Methods* (2nd Ed.). Oxford: Oxford University Press.
Bügel, K. & Buunk, B. (1996). Sex differences in foreign language text comprehension: The role of interests and prior knowledge. *Modern Language Journal*, 80, 15-31.
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.
Carifio, J. & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, 1150-1152.
Carroll, J. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, 1 (1), 131-51.
Carroll, J.B. & Sapon, S M. (1958). *The Modern Language Aptitude Test.* New York: Psychological Corp.
Carson, J. & Longhini, A. (2002). Focusing on learning styles and strategies: a diary study in an immersion setting. *Language Learning*, 52 (2), 401-438.
Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd Ed.). Boston, MA: Heinle.
Chall, J.S. & Dale, E. (1950). Familiarity of selected health terms. *Educational Research Bulletin*, 39, 197-206.
Chamot, A.U. (2004). Issues in language learning strategy research and teaching. *Electronic Journal of Foreign Language Teaching*, 1, 14-26.
Chamot, A.U., Kupper, L. & Impink-Hernandez, M. (1987). *A study of learning strategies in foreign language instruction: First year report.* Rosslyn, VA: InterAmerica Research Associates.

Chapelle, C.A. (1998). *Construct definition and validity inquiry in SLA research.* Cambridge: Cambridge University Press.
Chen, C. & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics,* 31, 693-713.
Chenoweth, A. & Hayes, J. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication,* 18, 80-98.
Chomsky, N. (1957). *Syntactic Structures.* The Hague, Netherlands: Mouton & Co.
Clark, J.L.D. (1981). Language. In T.S. Barrows, S.M. Ager, M.F. Bennett, H.I. Braun, J.L.D. Clark, L.G. Harris & S. F. Klein (Eds.), *College Students' Knowledge and Beliefs: A Survey of Global Understanding.* New Rochelle, NY: Change Magazine Press (pp. 25-35).
Clay, F., Bowers, J.S., Davis, C.J. & Hanley, D.A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology-Learning Memory and Cognition*, 33 (5), 970-976.
Coady, J. (1997). L2 vocabulary acquisition: a synthesis of research. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition*. New York: Cambridge University Press (pp. 225-237).
Cobb, T. (2000). One size fits all? Francophone learners and English vocabulary tests. *Canadian Modern Language Review,* 57 (2), 295-324.
Cobb, T. (2013). Frequency 2.0: Incorporating homoforms and multiword units into pedagogical frequency lists. In C. Bardel, C. Lindqvist & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis*. Eurosla Monographs Series, 2 (pp. 79-107).
Cobb, T. & Horst, M. (2004). Is there room for an AWL in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, acquisition, and testing.* Amsterdam: John Benjamins (pp. 15-38).
Cohen, A.D. & Shively, R.L. (2007). Acquisition of requests and apologies in Spanish and French: Impact of study abroad and strategy-building intervention. *The Modern Language Journal,* 91 (2), 189-212.
Coleman, J.A. & Chafer, T. (2010). Study abroad and the Internet: Physical and virtual context in an era of expanding telecommunications. *Frontiers: The Interdisciplinary Journal of Study Abroad,* 19, 151-167.
Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, 26 (2), 227-248.
Cowie, A. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2 (3), 223-235.
Cowie, A. (1988). Stable and creative aspects of vocabulary use. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching.* London: Longman (pp. 126-139).
Cowie, A. (1998). *Phraseology: Theory, analysis, and applications*. Oxford: Clarendon Press.
Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238.
Craik, F.I.M., & Lockhart. R.S. (1972). Levels of' processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*. II, 671-684.
Cresswell, J.W. (2003). *Research design: Qualitative, quantitative and mixed methods approaches* (2nd Ed.). Thousand Oaks, CA: Sage.
Creswell, J.W. & Plano Clark, V.L. (2007). *Designing and Conducting Mixed Methods Research.* Thousand Oaks, California: Sage.
Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied Linguistics*, 11, 183-199.

Cubillos, J.H., Chieffo, L. & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals,* 41, 157-185.
Cutting, J. & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally controlled idiom blends. *Memory and Cognition*, 25 (1), 57-71.
Daller, H., Milton, J. & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge.* Cambridge: Cambridge University Press.
Darwin, C.M. & Gray, L.S. (1999). Going after the Phrasal Verb: An Alternative Approach to Classification. *TESOL Quarterly,* 33 (1), 65-83.
Davidson, D.E. (2010). Language learning and study abroad. *Foreign Language Annals*, 43, 6-26.
Day, R.R. (1985). The use of the target language in context. In S. Gass and C. Madden (Eds.), *Language Input in Second Language Acquisition.* Rowley, MA: Newbury House Publishers (pp. 257- 271).
Day, R., Omura. C. & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7 (2), 541-551.
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*. Lexington, Mass.: DC Heath and Co. (pp. 104-119).
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*. Amsterdam: John Benjamins (pp. 313–326).
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., Bown, J. & Eggett, D. (2012). Japanese language proficiency, social networking, and language use during study abroad: Learners' perspectives. *Canadian Modern Language Review,* 68, 111-137.
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.
de Winter, J.C.F & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, 39 (4), 695-710.
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.
Dörnyei, Z. (2005). *Psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Erlbaum.
Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford, UK: Oxford University Press.
Dörnyei, Z. & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition.* Oxford, UK: Blackwell (pp. 589–630).
Dumay, N. & Gaskell, M.G. (2005). Do words go to sleep? Exploring consolidation of spoken forms through direct and indirect measures. *Behavioral and Brain Sciences, 28*, 69-70.
Dumay, N. & Gaskell, M.G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science,* 18, 35-39.
Dumay, N., Gaskell, M.G. & Feng, X. (2004). A day in the life of a spoken word. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates (pp. 339-344).
Duperron, L. (2006). Study abroad and the second language acquisition of tense and aspect in French: Is longer better? In S. Wilkinson (Ed.), *Insights from study abroad for language programs.* Boston: Heinle (pp. 47–51).
Dupuy, B. & Krashen, S. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning*, 4, 55-63.
Durrant, P. & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research,* 26 (2), 163-188.
Dwyer, M.M. (2004). Charting the impact of studying abroad. *International Education*, 4, 14-20.
Dyson, P. (1988). *The Year Abroad. Report for the Central Bureau for Educational Visits and Exchanges.* Oxford: Oxford University Language Teaching Centre.
Ehrman, M.E. & Leaver, B.L. (2003). Cognitive styles in the service of language learning. *System,* 31 (3), 393-415.
Ehrman, M. & Oxford, R. (1988). *Ants and grasshoppers, badgers and butterflies: Qualitative and quantitative exploration of adult language learning styles and strategies.* Paper presented at the Symposium on Research Perspectives on Adult Language Learning and Acquisition, Ohio State University, Columbus, OH.
Ehrman, M. & Oxford, R. (1990). Adult language learning styles and strategies in an intensive training setting. *Modern Language Journal,* 74 (3), 311-327.
Elgort, I. & Nation, I.S.P. (2010). Vocabulary learning in a second language: Familiar answers to new questions. In P. Seedhouse, S. Walsh and C. Jenks (Eds.), *Conceptualising 'learning' in applied linguistics*. Houndmills: Macmillan (pp. 89-104).
Ellis, N.C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description Acquisition and Pedagogy. Cambridge*: Cambridge University Press (pp. 122-139).
Ellis, N.C. (2008). Implicit and explicit knowledge about language. In J. Cenoz & N.H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd Ed.). Volume 6: Knowledge about language. New York: Springer (pp. 119-132).
Ellis, R. (1985). *Understanding second language acquisition.* Oxford, UK: Oxford University Press.

Ellis, R. (1994). *The study of second language acquisition* (1st Ed.). Oxford, UK: Oxford University Press.
Ellis, R. (1995). Modified input and the acquisition of word meanings. *Applied Linguistics,* 16 (4), 409-441.
Embassy English (2014). *Language year abroad long term English course* [Webpage]. Available at http://www.embassyenglish.com/courses/long-term-english/language-year-abroad [Last accessed 18 September 2014].
Erman, B. & Warren, B. (2000). The idiom principle and the open choice principle. *Text,* 20 (1), 29-62.
Eurolingua (2014). *Study abroad programs with Eurolingua* [Webpage]. Available at http://www.eurolingua.com/ [Last accessed 18 September 2014].
Færch, C., Haastrup, K. & Phillipson, R. (1984). *Learner language and language learning*. Clevedon, Avon, UK: Multilingual Matters.
Fan, M.Y. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31 (2), 105-119.
Fan, M.Y. (2003). Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *The Modern Language Journal*, 87, 222-241.
Félix-Brasdefer, J.C. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language Learning,* 54 (4), 587-653.
Firth, J.R. (1957). A synopsis of linguistic theory, 1930-1955'. In J.R. Firth et al. (Eds.), *Studies in Linguistic Analysis. Special volume of the Philological Society.* Oxford: Blackwell (pp. 1-32).
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*. London: Palgrave-Macmillan (pp. 91-106).
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.
Gardner, D. & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. TESOL Quarterly, 41, 339-359.
Gaskell, M.G. & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition,* 89, 105-132.
Gass, S.M. (2001). Innovations in second language research methods. *Annual Review of Applied Linguistics*, 21, 221-32.

Gass, S.M. & Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research.* Mahwah, New Jersey: Lawrence Erlbaum Associates.
Gass, S.M. & Selinker, L. (2001). *Second language acquisition: an introductory course* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
Geertz, C. (1973). *The interpretation of cultures: Selected essays.* New York: Basic Books.
Godfrey, L., Treacy, C. & Tarone, E. (2014). Change in French L2 writing in study abroad and domestic contexts. *Foreign Language Annals,* 47 (1), 48-65.
Goh, C.C.M., & Kwah, P.F. (1997). Chinese ESL students' learning strategies: A look at frequency, proficiency, and gender. *Hong Kong Journal of Applied Linguistics*, 2 (1), 39–53.
Goldstein, S.B. & Kim, R.I. (2006). Predictors of U. S. college students' participation in study abroad programs: A longitudinal study. *International Journal of Intercultural Relations*, 30, 507-521.
Golonka, E.M. (2006). Predictors Revised: Linguistic Knowledge and Metalinguistic Awareness in Second Language Gain in Russian. *Modern Language Journal,* 90 (4), 496-505.
Grainger, P.R. (1997). Language-learning strategies for learners of Japanese: investigating ethnicity. *Foreign Language Annals,* 30 (3), 378-385.
Green, J.M. & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency and gender. *TESOL Quarterly,* 29, 261-297.
Grenfell, M. & Harris, V. (2013). Learning to Learn Languages: the Differential Response of Learners to Strategy Instruction. *Curriculum Journal*, 24 (1), 121-152.
Grenfell, M. & Macaro, E. (2007). Claims and critiques. In A. Cohen & E. Macaro (Eds.), *Language Learner Strategies.* Oxford: Oxford University Press (pp. 9-28).
Griffiths, C. (2008). Strategies and good language learners. In C. Griffiths (Ed.), *Lessons from good language learners.* Cambridge, UK: Cambridge University Press (pp. 83–98).
Griffiths, C. (2013). *The strategy factor in successful language learning*. Bristol, UK: Multilingual Matters.
Gu, Y. (2002). Gender, academic major and vocabulary learning of Chinese EFL learners. *RELC Journal*, 33, 35-55.
Gu, Y. (2014). Personal email communication received 25 May 2014. Subject of message: *Putting you in touch*. Sent to jess.briggs@kellogg.ox.ac.uk.
Gu, Y. & Johnson, R. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning,* 46 (4), 643-679.
Guiraud, P. (1954). *Les Charactères Statistiques du Vocabulaire. Essai de méthodologie*. Paris: Presses Universitaires de France.
Guntermann, G. (1995). The Peace Corps experience: Language learning in training and in the field. In B.F. Freed (Ed.), *Second language acquisition in a study abroad context.* Amsterdam: John Benjamins (pp.149-169).
Gyllstad, H. (2007). *Testing English Collocations - Developing Receptive Tests for Use with Advanced Swedish Learners.* Lund University: Media-Tryck.
Halliday, M.A.K. (1961). Categories of the theory of grammar. *Word*, 17, 241-292.
Halliday, M.A.K. (1966). Lexis as a linguistic level. In C.E. Bazell, J.C. Catford, M.A.K. Halliday & R.H. Robbins (Eds.), *In memory of F.R. Firth.* London: Longman (pp. 150-161).

Harrison, J. K. & Voelker, E. (2008). Two personality variables and the cross-cultural adjustment of study abroad students. *Frontiers: The Interdisciplinary Journal of Study Abroad,* 17, 69-87.
Hassall, T. (2013). Pragmatic development during short-term study abroad: The case of address terms in Indonesian. *Journal of Pragmatics*, 55, 1-17.
Hatasa, Y. & Soeda, E. (2000). Writing strategies revisited: A case of non-cognate L2. In B. Swierzbin, F. Morris, M. Anderson, C. Klee and E. Tarone (Eds.), *Social and cognitive factors in second language acquisition.* Sommerville: Cascadilla Press (pp. 375-96).
Henriksen, B. (1996). *Semantisation, retention, and accessibility: Key concepts in vocabulary learning.* Paper presented at the AILA Congress, Jyvaskyla, Finland.
Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21 (2), 303-317.
Henriksen, B. (2013). Research on L2 learners' collocational competence and development – a progress report. In C. Bardel, C. Lindqvist & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: new perspective on assessment and corpus analysis.* Online: Eurosla Monographs Series 2 (pp. 29-56).
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.
Hong-Nam, K. & Leavell, A. (2006). Language learning strategy use of ESL students in an intensive English learning context. *System*, 34, 399-415.
Howarth, P. (1996). *Phraseology in English Academic Writing : Some implications for language learning and dictionary making.* Berlin: De Gruyter.
Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics,* 19 (1), 24-44.
Hsueh-chao, M.H. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language,* 13 (1), 403-430.
Huebner, T. (1995). The effects of overseas language programs. In B.F. Freed (Ed.), *Second language acquisition in a study abroad context.* Philadelphia, PA: John Benjamins Publishing Company (pp. 172-193).
Hu, M. & Nation, I.S.P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13 (1), 403-30.
Hulstijn, J.H. (1992). Retention of inferred and given word meanings: Experiments in incidental learning. In P.J.L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics.* London: Macmillan (p.113-125).
Hulstijn, J.H. (2003). Incidental and intentional learning. In C. Doughty & M.H. Long (Eds.), *The handbook of second language research.* London: Blackwell (pp. 349-381).
Hulstijn, J.H., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning,* 51, 539-558.
Hulstijn, J.H. & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition*, 11 (3), 241-255.
Hurd, S. & Lewis, T. (2008). *Language Learning Strategies in Independent Settings.* Clevedon, UK: Multilingual Matters.

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.
Institute of International Education (2013). *Open Doors Report on International Educational Exchange.* Available at http://www.iie.org/Research-and-Publications/Open-Doors [Last accessed 01 August 2014].
Isabelli, C. (2004). Study abroad for advanced foreign language majors: Optimal duration for developing complex structures In H. Byrnes & H. Maxim (Eds.), *Advanced Foreign Language Learning, A Challenge to College Programs.* Canada: Heinle (pp. 114-130).
Isabelli, C. (2007). Development of the Spanish subjunctive by advanced learners: Study abroad followed by at-home instruction. *Foreign Language Annals,* 4 (2), 330-41.
Isabelli-García, C. (2006). Study abroad social networks, motivation, and attitudes: Implications for second language acquisition. In M.A. Dufon & E. Churchill (Eds.), *Language learners in study abroad contexts.* Cleveland, OH: Multilingual Matters (pp. 231-258).
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.
Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38, 1212-1218.
Janssens, V. (1999). Over 'slapen' en 'snurken' en de hulp van de context hierbij. *ANBF-nieuwsbrief*, 4, 29–45.
Johnson, J.S. & Newport, E.L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.
Johnson, B. & Turner, L.A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioural research.* Thousand Oaks, CA: Sage Publications (pp. 297-320).
Juan-Garau, M. & Pérez-Vidal, C. (2007). The effect of context and contact on oral communication in students who go on a stay abroad. *EUROSLA Yearbook 2007*. Amsterdam: Walter Benjamins.
Judd, C.M. & McClelland, G.H. (1989). *Data analysis: A model comparison approach.* San Diego, CA.: Harcourt Brace Jovanovich.
Kaiser, H. (1970). A second generation Little Jiffy. *Psychometrica,* 35, 401-415.
Kaiser, H. (1974). An index of factorial simplicity. *Psychometrica*, 39, 31-36.
Kaylani, C. (1996). The influence of gender and motivation on EFL learning strategy use in Jordan. In Oxford, R.L. (Ed.), *Language learning strategies around the world: Crosscultural perspectives.* Manoa: University of Hawaii Press (pp.75-88).
Kaypak, E. & Ortaçtepe, D. (2014). Language learner beliefs and study abroad: A study on English as a lingua franca (ELF). *System,* 42, 355-367.
Keatley, C., Chamot, A., Spokane, A. & Greenstreet, S. (2004). Learning strategies of students of Arabic. *The Language Resource*, 8 (4). Available at http://www.nclrc.org/nectfl041s.pdf [Last accessed 15 August 2014].
Khalil, A. (2005). Assessment of language learning strategies used by Palestinian EFL learners. *Foreign Language Annals,* 38, 108-119.
Kinginger, C. (2008). Language learning in study abroad: Case studies of Americans in France. *Modern Language Journal*, 92, Monograph.
Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research.* London: Palgrave Macmillan.

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.
Krashen, S.D. (1987). *Principles and practices in second language acquisition*. New York: Prentice-Hall.
Krashen, S.D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal,* 73, 440-464.
Kline, R.R. (1998). Literacy and language learning in a study abroad context. *Frontiers, the Interdisciplinary Journal of Study Abroad,* 4, 139-166.
Knoke, D. and Yang, S. (2008). Social network analysis. Thousand Oaks, CA: Sage Publications.
Kojic-Sabo, I. & Lightbown, P.M. (1999). Students' approaches to vocabulary learning and their relationship to success. *Modern Language Journal,* 83 (2), 176-192.
Koya, T. (2005). *The acquisition of basic collocations by Japanese learners of English*. Unpublished doctoral thesis. Waseda University, Japan.
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.
Lafford, B. & Collentine, J. (2006) The effects of study abroad and classroom contexts on the acquisition of Spanish as a second language: From research to application. In B. Lafford & R. Salaberry (Eds.), *Spanish Second Language Acquisition: From Research Findings to Teaching Applications*. Washington, DC: Georgetown University Press (pp. 103-126).
Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33, 159-174.
Lapkin, S., Hart, D. & Swain, M. (1995). A Canadian Interprovincial Exchange: Evaluating the Linguistic Impact of a Three-Month Stay in Quebec. In B.F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context.* Amsterdam and Philadephia: John Benjamins (pp. 67-94).
Laufer, B. (1988). The concept of 'synforms' (similar lexical forms) in L2 vocabulary acquisition. *Language and Education,* 2, 113-132.
Laufer, B. (1989). What percentage of lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: From Human Thinking to Thinking Machines*. Multilingual Matters (pp. 316-23).
Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and Applied Linguistics*. Basingstoke and London: Macmillan (pp. 126-132).
Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 12, 255-271.
Laufer, B. (2005). Lexical Frequency Profiles: From Monte Carlo to the Real World. A response to Meara. *Applied Linguistics*, 26, 581-587.
Laufer, B. & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning*, 54, 469-523.
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.
Laufer, B. & Waldman, T. (2011). Verb-noun collocations in second language writing: a corpus analysis of learners' English. *Language Learning*, 61 (2), 647-672.
Lawson, M.J. & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning,* 46, 101-135.
LearnEnglishGuide (2014). *Why learn English abroad* [Webpage]? Available at http://www.learnenglishguide.com/abroad/ [Last accessed 17 September 2014].
Leech, G., Rayson, P. & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus.* London: Longman.
Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning,* 40, 387-412.
Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward.* Hove, England: Language Teaching Publications.
Liao, Y. & Fukuya, Y.J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, 54 (2), 193-226.
Lincoln, Y.S. & Guba, E.G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.
Liyanage, I. (2004). *An exploration of language learning strategies and learner variables of Sri Lankan Learners of English as a second language with special reference to their personality types.* Unpublished doctoral thesis. Griffith University, Queensland, Australia.
Liyanage, I. & Bartlett, B.J. (2012). Gender and language learning strategies: looking beyond the categories. *The Language Learning Journal*, 40 (2), 237-253.
Llanes, À. & Muñoz, C. (2009). A short stay abroad: does it make a difference? *System,* 37 (3), 353-365.
Llanes, À. & Muñoz, C. (2013). Age effects in a study abroad context: children and adults studying abroad and at home. *Language Learning,* 63 (1), 63-90.
Macaro, E. (2001) *Learning Strategies in second and foreign language classrooms.* London: Continuum.
Macaro, E. (2003). *Teaching and Learning a Second Language: a guide to current research and its applications.* London: Continuum.
Macaro, E. (2004). *Fourteen Features of a Language Learner Strategy.* Working Paper No. 4. Oxford: Department of Educational Studies, University of Oxford.
Macaro, E. (2006). Strategies for language learning and for language use: revising the theoretical framework. *Modern Language Journal*, 90 (3), 320-337.
Magnan, S.S. & Back, M. (2007). Social interaction and linguistic gain during study abroad. *Foreign Language Annals*, 40, 43-61.
Magogwe, J. & Oliver, R. (2007). The relationship between language learning strategies, proficiency, age and self-efficacy beliefs: a study of language learners in Botswana. *System*, 35, 338-352.
Malvern, D.D. & Richards, B.J. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing,* 19, 85-104.
Martinez, R. & Murphy, V.A. (2011). The effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, 45, 267-290.
Martinez, R. & Schmitt, N. (2012). A Phrasal Expressions List. *Applied Linguistics,* 33 (3), 299-320.
Martinsen, R.A. (2011). Predicting changes in cultural sensitivity among students of Spanish during short-term study abroad. *Hispania*, 94, 121-141.

McCarthy, M. (1990). *Vocabulary.* Oxford: Oxford University Press.
McDonough, S.M. (1995). *Strategy and Skill in Learning a Foreign Language.* London: Edward Arnold.
McLaughlin, J., Osterhout, L. & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience,* 7, 703-704.
Meara, P.M. (1994). The Year Abroad and Its Effects. *Language Learning Journal,* 10, 32-38.
Meara, P.M. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (Eds.), *Performance and competence in second language acquisition.* Cambridge: Cambridge University Press (pp. 35-53).
Meara, P.M. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press (pp. 109-121).
Meara, P.M. & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospects,* 16 (3), 5-19.
Meara, P.M. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing,* 4, 142-145.
Meara, P.M. & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System,* 28, 19-30.
Meara, P.M. & Jones, G. (1990). *The Eurocentres Vocabulary Size Tests: 10K*. Zurich: Eurocentres.
Meara, P.M. & Milton, J. (2003). *X_Lex, the Swansea levels test*. Newbury: Express.
Meara, P.M. & Miralpeix, I. (2006). *Y_Lex* [Computer Program]. Swansea, UK: Lognostics.
Meara, P.M. & Wolter, B. (2004). V_Links, beyond vocabulary depth. *Angles on the English Speaking World*, 4, 85-96.
Medina-Lopez-Portillo, A. (2004). Intercultural learning assessment: The link between program duration and the development of intercultural sensitivity. *Frontiers: The Interdisciplinary Journal of Study Abroad,* 10, 179-200.
Melka Teichroew, F.J. (1982). Receptive vs. productive vocabulary: a survey. *Interlanguage Studies Bulletin*, 6, (2), 5-33.
Miller, L. & Ginsberg, R. (1995). Folklinguistic Theories of Language Learning. In B.F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context.* Amsterdam and Philadelphia: John Benjamins Publishing Company (pp. 293-315).
Milton, J. (2006). Language Lite: Learning French vocabulary in school. *Journal of French Language Studies*, 16 (2), 187-205.
Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
Milton, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4 (3), 151-172.
Milton, J. & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review,* 63 (1), 127-147.
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.
Milton, J. & Riordan, O. (2006). Level and script effects in the phonological and orthographic vocabulary size of Arabic and Farsi speakers. In P. Davidson, C.

Coombe, D. Lloyd & D. Palfreyman (Eds.), *Teaching and Learning Vocabulary in Another Language.* UAE: TESOL Arabia (pp. 122-133).
Milton J., Wade, J. & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse & M. Torreblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning.* Bristol: Multilingual Matters (pp. 83-98).
Mochida, A. & Harrington, M.W. (2006). The Yes-No test as a measure of receptive vocabulary knowledge. *Language Testing,* 23, 73-98.
Moir, J. & Nation, I.S.P. (2002). Learners' use of strategies for effective vocabulary learning. *Prospects,* 17 (1), 15-35.
Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press (pp. 40-63).
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*. Clevedon, England: Multilingual Matters (pp. 241-63).
Mora, J.C. (2014). The role of onset level on L2 perceptual phonological development after formal instruction and study abroad. In C. Pérez-Vidal (Ed.) *Language Acquisition in Study Abroad and Formal Instruction Contexts*. Amsterdam: John Benjamins (pp. 167-194).
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.
Moreno Espinosa, S. (2009). Young learners' L2 word association responses in two different learning contexts. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe.* Clevedon: Multilingual Matters (pp. 93-111).
Morgan, D.L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research,* 1 (1), 48-76.
Mori, S. (2007). Language learning strategy use by learners of Japanese at different levels. *Paper of the Japanese Language Teaching Association in Honor of Professor Fumiko Koide*, 15, 87-104.
Morse, J.M. (2003). Principles of mixed methods and multimethod research design. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioural research.* Thousand Oaks, CA: Sage (pp. 189-208).
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.
Myers, M. (2000). *Qualitative research and the generalizability question: Standing firm with Proteus*. The Qualitative Report [Electronic], 4 (3/4). Available from: http://www.nova.edu/ssss/QR/QR4-3/myers.html [Last accessed 22 June 2010].
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. (1983). Testing and teaching vocabulary. *Guidelines*, 5 (1), 12-24.
Nation, I.S.P. (1990). *Teaching and learning vocabulary.* New York: Newbury House.
Nation, I.S.P. (2001). *Learning vocabulary in another language.* Cambridge, UK: Cambridge University Press.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63* (1), 59-81.
Nation, I.S.P. (2007) Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton & J. Treffers-Daller (eds.), *Modelling and Assessing Vocabulary Knowledge.* Cambridge: Cambridge University Press (pp. 35-43).
Nation, I.S.P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher,* 31 (7), 9-13.
Nattinger, J.R. & DeCarrico, J.S. (1992). *Lexical Phrases and Language Teaching.* Oxford: Oxford University Press.
Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics,* 24 (2), 223-242.
Nesselhauf, N. (2005). *Collocations in a Learner Corpus.* Amsterdam and Philadelphia: Benjamins.
Nguyen, L. & Nation, I.S.P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal,* 42 (1), 86-99.
Norris, B. (2012). *Global Market 2012* [Online resource]. Hot House Media Ltd.: Study Travel Magazine. Available at http://www.hothousemedia.com/ltm/ltmbackissues/dec13web/dec13specreport.html [Last accessed 12 September 2014].
Nurweni, A. & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes,* 18, 161-175.
Olivares-Cuhat, G. (2002). Learning strategies and achievement in the Spanish writing classroom: A case study. *Foreign Language Annals,* 35 (5), 561-570.
O'Malley, J.M. & Chamot, A.U. (1990) *Learning Strategies in Second Language Acquisition.* Cambridge: Cambridge University Press.
O'Malley, J.M. & Chamot, A.U. (1993). Learner characteristics in second language acquisition. In A. Omaggio Hadley (Ed.), *Research in language learning: Principles, processes, and prospects* (ACTFL Foreign Language Education Series, Vol. 24). Lincolnwood, IL: National Textbook (pp. 96-123).
O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Kupper, L. & Russo, R. P. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly,* 19, 557-584.
Orr, J.M., Sackett, P.R. & DuBois, C.L.Z. (1991). Outlier detection and treatment in I/O Psychology: A survey of researcher beliefs and an empirical illustration. *Personnel Psychology*, 44, 473-486.
Osborne, J.W. (2002). Notes on the use of data transformations. *Practical Assessment, Research, and Evaluation,* 8. Available at http://ericae.net/pare/getvn.asp?v=8&n=6 [Last accessed 24 May 2014].
Osborne, J.W. & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). Practical *Assessment, Research & Evaluation*, 9 (6). Available at http://PAREonline.net/getvn.asp?v=9&n=6 [Last accessed 24 May 2014].
Oxford, R. (1993). Instructional implications of gender differences in language learning styles and strategies. *Applied Language Learning*, 4 (1&2), 65-94.
Oxford, R.L. (1990). *Language Learning Strategies. What Every Teacher Should Know.* Heinle, Boston.
Oxford, R.L. (2011). *Teaching and researching language learning strategies.* Harlow: Longman.

Oxford, R. & Burry-Stock, J. (1995). Assessing the use of language learning strategies worldwide with the ESL/EFL version of the strategy inventory for language learning (SILL). *System,* 23, 1-23.
Oxford, R. & Ehrman, M. (1995). Adults' language learning strategies in an intensive foreign language program in the United States. *System,* 23, 359-386.
Oxford, R., Lavine, R., Felkins, G., Holloway, M. & Saleh, A. (1996). Telling their Stories: Language Students Use Diaries and Recollection. In Oxford, R. (Ed.), *Language Learning Strategies around the World: Cross-Cultural Perspectives* (Technical Report 13). Second Language Teaching and Curriculum Center: University of Hawai'i Press (pp. 19-34).
Oxford, R. & Nyikos, M. (1989). Variables affecting choice of language learning strategies by university students. *Modern Language Journal,* 73 (3), 291-300.
Oxford, R. & Schramm, K. (2007). Bridging the gap between psychological and sociocultural perspectives on L2 learner strategies. In A.D. Cohen & E. Macaro (Eds.), *Language Learner Strategies*. Oxford: Oxford University Press (pp. 47-68).
Ozánska-Ponikwia, K. & Dewaele, J. M. (2012). Personality and L2 use: The advantage of being openminded and self-confident in an immigration context. *EUROSLA Yearbook,* 12, 112–134.
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.
Palmberg, R. (1987). Patterns of vocabulary development in foreign language learners. *Studies in Second Language Acquisition*, 9, 202-221.
Palmer, H.E (1917). *The Scientific Study and Teaching of Languages.* London: Harrap.
Paribakht, T.S. & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal,* 11 (1), 9-29.
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.
Park, G. (1997). Language learning strategies and English proficiency in Korean university students. *Foreign Language Annals*, 30 (2), 211-221.
Peacock, M. & Ho, B. (2003). Student language learning strategies across eight disciplines. *International Journal of Applied Linguistics,* 13 (2), 179-200.
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.* Sydney: Thomson Heinle (pp. 99-115).
Pellicer-Sánchez, A. & Schmitt, N. (2012). Scoring Yes-No vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, 29 (4), 489-509.
Pérez-Vidal, C. & Juan-Garau, M. (2009). The effect of Study Abroad (SA) on written performance. *Eurosla Yearbook 9*. John Benjamins Publishing Company.
Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.
Pigada, M. & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language,* 18, 1-28.
Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers*, 4, 37-45.

Pirker, C. (2009). *Statistical noise or valuable information the role of extreme cases in marketing research.* Wiesbaden: Gabler Verlag/GWV Fachverlage GmbH.
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.*
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*. Amsterdam: John Benjamins (pp. 271–291).
Punch, K.F. (2009). *Introduction to research methods in education.* London: Sage.
Qian, D.D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56 (2), 282-307.
Qian, D.D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning,* 52 (3), 513-36.
Qian, D.D. (2004). Second language lexical inferencing: Preferences, perceptions, and practices. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing.* Amsterdam: Benjamins, (pp. 155-169).
Ramirez, A.G. (1986). Language Learning Strategies used by adolescents studying French in New York schools. *Foreign Language Annals*, 19 (2), 131-141.
Razi, O. & Grenfell, M. (2012). The impact of Linguistic knowledge on Learner Strategy Deployment. *Procedia Social and Behavioral Sciences,* 47, 818-822.
Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal,* 19 (2), 12-25.
Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10, 355-371.
Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in Language Assessment*. Mahwah, NJ: Erlbaum (pp. 41-60).
Read, J. (2000). *Assessing vocabulary.* Cambridge, UK: Cambridge University Press.
Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing.* Amsterdam: Benjamins, (pp. 209-227).
Read, J. (2007). *Extending the Yes/No format as a measure of vocabulary knowledge*. Paper given at the Annual Congress of the Applied Linguistics Association of Australia (ALAA), Wollongong, July.
Read, J. & Chapelle, C.A. (2001). A framework for second language vocabulary assessment. *Language Testing,* 18, 1-32.
Reynolds-Case, A. (2013). The value of short-term study abroad: An increase in students' cultural and pragmatic competency. *Foreign Language Annals*, 46 (2), 311-322.
Richards, B.J. & Malvern, D.D. (2007). Validity and threats to the validity of vocabulary assessment. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge.* Cambridge: Cambridge University Press (pp. 79-92).
Richards, J.C. (1976). The role of vocabulary teaching. *TESOL Quarterly,* 10 (1), 77-89.
Rigney, J. (1978). Learning strategies: A theoretical perspective. In H. F. Jr. O'Neil (Ed.), *Learning Strategies.* New York: Academic Press.

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.
Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly*, 9, 41-51.
Rubin, J. (1987). Learner strategies: Theoretical assumptions, research history and typology. In A. L. Wenden & J. Rubin (Eds.), *Learner strategies in language learning*. Englewood Cliffs, NJ: Prentice-Hall (pp. 15-30).
Ryle, G. (1949). *Concept of the mind*. London: Hutchinson and Company.
Sandelowski, M. (2003). Tables or tableaux? The challenges of writing and reading mixed methods studies. In A. Tashakkori and C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, CA: Sage (pp. 321-350).
Sanaoui, R (1992). *Vocabulary learning and teaching in French as a second language classrooms.* Unpublished doctoral dissertation. Toronto: University of Toronto, OISE.
Sasaki, M. (2004). A multiple data analysis of the 3.5-year development of EFL student writers. *Language Learning*, 54 (3), 525-582.
Sasaki, M. (2007) Effects of Study abroad Experiences on EFL Writers: A Multiple-Data Analysis. *The Modern Language Journal*, 91 (4), 602-620.
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*. Clevendon: Multilingual Matters (pp. 49-76).
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction.* Cambridge, UK: Cambridge University Press (pp. 3-32).
Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press (pp. 199-227).
Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning,* 48 (2), 281-317.
Schmitt, N. (2000). *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press.
Schmitt, N. (Ed.). (2004). *Formulaic Sequences: Acquisition, Processing, and Use.* Amsterdam: John Benjamins Press.
Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research,* 12, 329–363.
Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Press.
Schmitt, N. & Carter, R. (2004). Formulaic sequences in action. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use.* Amsterdam: John Benjamins (pp. 1-22).
Schmitt, N. & McCarthy, M. (1997). *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press.
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.
Schoonen, R. (2010). *The development of lexical proficiency knowledge and skill*. Paper presented at the Copenhagen Symposium on Approaches to the Lexicon, Copenhagen Business School on 8-10 December 2010. Available at https://conference.cbs.dk/index.php/lexicon/lexicon/schedConf/presentations [Last accessed 14 July 2014].

Schouten-Van Parreren, C. (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts? V*ocabulary Acquisition AILA Review*, 24 (6), 75-85.
Scott, J. (2000). *Social Network Analysis: A Handbook.* London: Sage.
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.
Segalowitz, N. & Ryder, A. (2006). *Montreal Index of Linguistic Integration (MILI).* Unpublished questionnaire, Concordia University, Montreal. Quebec, Canada: Concordia University.
Seliger, H. (1977). Does practice make perfect? A study of interaction patterns and L2 competence. *Language Learning,* 27, 263-278.
Seliger, H. (1983). The language learner as linguist: Of metaphors and realities. *Applied Linguistics,* 4, 179-191.
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.
Shenton, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information,* 22, 63-75.
Sheorey, R. (1999). An examination of language learning strategy use in the setting of an indigenized variety of English. *System,* 27, 173-190.
Shin, D. & Nation, I.S.P. (2008). Beyond single words: the most frequent collocations in spoken English. *ELT Journal*, 62 (4), 339-348.
Shively, R.L. (2013). Learning to be funny in Spanish during study abroad: L2 humor development. *The Modern Language Journal*, 97 (4), 930-946.
Shmais, W.A. (2003). Language learning strategy use in Palestine. *TESL-EJ*, 7 (2), A-3.
Sinclair, J.M. (1987). *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. London: Collins ELT.
Sinclair, J.M. (1991). *Corpus, concordance, collocation: Describing English language.* Oxford: Oxford University Press.
Sonbul, S. & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 63 (1), 121-159.
Sparks, R.L. & Ganschow, L. (1993). The impact of native language learning problems on foreign language learning: case study illustrations of the linguistic coding deficit hypothesis. *Modern Language Journal,* 77, 58-74.
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.
Stern, H.H. (1975). What can we learn from the good language learner? *Canadian Modern Language Review,* 31, 304-318.
Stern, H.H. (1992). *Issues and Options in Language Teaching*. Oxford: OUP.
Stevens, G. (1999). Age at immigration and second language proficiency among foreign born adults. *Language in Society*, 28, 555-578.

Stevick, E.W. (1990). Research on what? Some terminology. *Modern Language Journal,* 74, 143-153.
Stewart, J. & White, D.A. (2011). Estimating guessing effects on the Vocabulary Levels Test for differing degrees of word knowledge. *TESOL Quarterly*, 45, 370-380.
Stubbe, R., Stewart, J. & Pritchard, T. (2010). Examining the Effects of Pseudowords in Yes/No Vocabulary Tests for Low Level Learners. *Kyushu Sangyo University Language Education and Research Center Journal,* 5, 5-23.
Sunderman, G. & Kroll, J.F. (2009). When the study abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics,* 30, 79-99.
Swain, M. (1985). Communicative competence: Some roles of comprehensive input and comprehensible output in its development. In S.M. Gass & C. Madden (Eds.), *Input in second language acquisition.* Cambridge, MA: Newbury House (pp. 235–256).
Swain, M. (1993). The output hypothesis: just speaking and writing aren't enough. *Canadian Modern Language Review*, 50 (1), 58-64.
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.
Tabachnick, B.G. & Fidell, L.S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.
Tabachnick, B.G. & Fidell, L.S. (2001). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.
Taguchi, N. (2008). Cognition, language contact, and the development of pragmatic comprehension in a study abroad context. *Language Learning*, 58 (1), 33-71.
Takeuchi, O., Ikeda, M. & Mizumoto, A. (2012). The cerebral basis for language learner strategies: A near-infrared spectroscopy study. *Reading in a Foreign Language*, 24 (2), 136-157.
Takeuchi, O., Griffiths, C. & Coyle, D. (2007). Applying strategies to contexts: the role of individual, situational, and group differences. In A.D. Cohen & E. Macaro (Eds.), *Language Learning Strategies.* Oxford: Oxford University Press (pp. 68-92).
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.
Tanaka, K. & Ellis, R. (2003). Study Abroad, Language Proficiency, and Learner Beliefs About Language Learning. *JALT Journal*, 25, 63–85.
Tashakkori, A. & Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, CA: Sage.
Thorndike, E.L. and Lorge, I. (1944). *The teacher's word book of 30,000 words.* New York: Teachers College Press.
Tragant, E., Thompson, M.S. & Victori, M. (2013). Understanding foreign language learning strategies: A validation study. *System,* 41, 95-108.
Tragant, E. & Victori, M. (2012). Language learning strategies, course grades, and age in EFL secondary school learners. *Language Awareness,* 21, 293-308.
Tréville, M.C. (1988). Faut-il enseigner le vocabulaire dans la langue seconde? In Leblanc, R., Compain, R., Duquette, L. & Séguin, H. (Eds.), *L'enseignement des langues secondes aux adultes: recherches et pratiques.* Ottawa: Presses de l'Université d'Ottawa.

Tseng, W.T., Dörnyei, Z. & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics,* 27 (1), 78-102.
Tuckman, B.W. & Harper, B.E. (2012). *Conducting Educational Research*. Lanham, Maryland: Rowman and Littlefield.
Twombly, S.B. (1995). Piropos and friendship: Gender and culture clash in study abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 1, 1-27.
Vande Berg, M., Connor-Linton, J, & Paige, R.M. (2009). The Georgetown Consortium Project: Interventions for student learning abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad,* 18, 1-75.
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.
Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics,* 22 (2), 217-218.
Vidal, K. (2003). Academic listening: a source of vocabulary acquisition? *Applied Linguistics,* 24 (1), 56-89.
Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning,* 61, 219-258.
Vives Boix, G. (1995). *The development of a measure of lexical organisation: the Association Vocabulary Test.* Unpublished doctoral thesis. University of Wales, Swansea.
Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Assessment Quarterly*, 9 (2), 172-185.
Wang, A.Y., Thomas, M.H., Inzana, C.M. & Primicerio, L.J. (1993). Long-term retention under conditions of intentional learning and the keyword mnemonic. *Bulletin of the Psychology Society*, 31, 545-547.
Waring, R. & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language,* 15, 1–27.
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.
Webb, S. (2007). The effects of synonymy on vocabulary learning. *Reading in a Foreign Language*, 19, 120-36.
Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30, 79-95.
Webb, S. & Rodgers, M.P.H. (2009). The lexical coverage of movies. *Applied Linguistics,* 30, 407-427.
Weinstein, C.E. & Mayer, R.E. (1986). The teaching of learning strategies. In M. Wittrock (Ed.), *Handbook of Research on Teaching*. New York: Macmillan (pp. 315-327).
Wesche, M. & Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53 (1), 13-40.
West, M. (1953). *A general service list of English words.* London: Longman.
Wharton, G. (2000). Language learning strategy use of bilingual foreign language learners in Singapore. *Language Learning,* 50 (2), 203-243.
White, C., Schramm, K. & Chamot, A. (2007). Research methods in strategy research: re-examining the toolbox. In A. D. Cohen & E. Macaro (Eds.), *Language learner*

strategies: 30 years of research and practice. Oxford: Oxford University Press (pp. 93-116).
Wilkins, D.A. (1972). *Linguistics and Language Teaching.* London: Edward Arnold.
Wilkinson, S. (1998). Study abroad from the participants' perspective: A challenge to common beliefs. *Foreign Language Annals*, 31 (1), 23-39.
Wilkinson, S. (2002). The omnipresent classroom during summer study abroad: American students in conversation with their French hosts. *Modern Language Journal,* 86 (2), 157-173.
Williams, R. (1980). Lexical familiarization in content area textbooks. In L.J. Chapman, (Ed.), *The Reader and the Text.* London: Heinemann.
Williams, R. (1985). Teaching vocabulary recognition strategies in ESP reading. *ESP Journal,* 4 (2): 121-131.
Winne, P.H. & Perry, N.E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation.* Orlando, FL: Academic Press (pp. 531–566).
Wolter, B. (2005). *V_Links: A New Approach to Assessing Depth of Word Knowledge.* Unpublished doctoral thesis. University of Wales, Swansea.
Wolter, B. & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing: A Comparison of Congruent and Incongruent Collocations. *Studies in Second Language Acquisition,* 35, 451-482.
Wong Fillmore, L. (1979). Individual Differences in Second Language Acquisition. In C. J. Fillmore, D. Kempler & W. S. Y. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior*. New York: Academic Press.
Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review*, 63, 13-33.
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.
Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.
Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: OUP.
Xing, P. & Fulcher, G. (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 Gain. *Hispania*, 81 (4), 898-913.
Yang, E-Y. & Jiménez, R.T. (2011). A Sociocultural Perspective on Second Language Learner Strategies: Focus on the Impact of Social Context. *Theory Into Practice*, 50, 141-148.
Young, D.J. & Oxford, R. (1997). A gender-related analysis of strategies used to process written input in the native language and a foreign language. *Applied Language Learning*, 8 (1), 1-20.
Zhang. B. & Li, C. (2011). Classification of L2 vocabulary learning strategies: Evidence from exploratory and confirmatory factor analyses. *RELC Journal,* 42, 141-154.

Appendix A: Study information letter

UNIVERSITY OF OXFORD
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk

A Study of the Relationships between Informal Second Language Contact, Strategic Behaviour and Vocabulary Gain in a Study abroad Context.

Hello! My name is Jess and I'm a graduate student at the Department of Education, University of Oxford. I'm interested in how you learn new English vocabulary outside of the classroom. I'd like to invite you to participate in my study called **'A Study of the Relationships between Informal Second Language Contact, Strategic Behaviour and Vocabulary Gain in a Study abroad Context'.** The study aims to find out what kind of out-of-class situations and behaviours are the most helpful for learning new vocabulary.

You are invited to be a part of this study because your UK language school has agreed to help with my research. All adult students at your school who are pre-intermediate level or above are invited to participate.

If you agree to be in the study, you will be invited to do 3 things:

1) Take a vocabulary test at the beginning and end of your time in England. The test is very short – it only takes 40 minutes in total.

2) Complete a questionnaire about what you do outside of the classroom in England. This will take a maximum of 40 minutes to complete and your answers will be kept a secret.

3) Be interviewed by me. The interview will ask you about your thoughts and behaviour when you see/hear new words in out-of-class situations. I will use a computer to show you the situations. It will last for about 40 minutes and, if you agree, I will audio-record the interview so I can better remember your answers!

I'll come to your language school to collect this information from you, and I will contact you before I come to make sure you have time to see me.

I will keep any information that you give me on a secure, password-controlled computer. Only I and my supervisor, Professor Macaro, will have access to the information you give. The information will only be used for the purpose of writing about my study. I won't use your real name in anything I write, and your course and UK language school will not be named.

If you get involved, your input could help researchers, teachers and students of English to understand better how learners of English can improve their vocabulary during their stay England. Participation is completely voluntary- if you change your mind, you can leave the study at any time.

This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.

I'd really like you to be a part of my study. If you have any more questions, please contact me by phone on +44(0)7805 305 975 or email jess.briggs@education.ox.ac.uk or my supervisor on ernesto.macaro@education.ox.ac.uk . If your query is unresolved, please contact Dr Lars Malmberg, Chair of the Departmental Research Ethics Committee by email: lars-erik.malmberg@education.ox.ac.uk or telephone: 01865 274047.

Appendix B: CUREC approval confirmation

From: **Research Office** research.office@education.ox.ac.uk
Subject: CUREC Application Approval
Date: 15 February 2012 16:33
To: Jess Briggs jess.briggs@kellogg.ox.ac.uk
Cc: Ernesto Macaro ernesto.macaro@education.ox.ac.uk

15/02/2012

Dear Jessica Briggs,

Application Approval

Title: "Close Encounters of the Lexical Kind: A Study of the Relationships between Informal Second Language Contact, Strategic Behaviour and Vocabulary Gain in a Study-Abroad Context."

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly approval has been granted.

Should there be any subsequent changes to the project, which raise ethical issues not covered in the original application, you should submit details to DREC for consideration.

Good luck with your research study.

Yours sincerely,

Justina Kurkova

Research Office Assistant

Appendix C: Consent form

UNIVERSITY OF OXFORD

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk

A Study of the Relationships between Informal Second Language Contact, Strategic Behaviour and Vocabulary Gain in a Study abroad Context.

This study aims to investigate the relationships between out-of-class contact with English, employment of vocabulary acquisition strategies and vocabulary gain in a study abroad context. This is a study undertaken by Jess Briggs, a doctoral student at the Department of Education, University of Oxford. You can contact this researcher using the address above, or via telephone number +44 (0) 7805 305975 or email jess.briggs@education.ox.ac.uk .

Please read the following text carefully:

- I have read and understood the information about this study as provided by the information sheet and I have had the opportunity to ask questions and get satisfactory answers about this study.
- I understand that I can withdraw from the study without any consequence at any time simply by informing the researcher of my decision.
- I understand who will have access to the identifying information provided and what will happen to the data at the end of this project.
- I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.

I agree to participate in this study.

Signed: ….....

Print name (block capitals): ….....

Date ….....

Signed by researcher: ….....

Print name (block capitals): JESS BRIGGS

Date ….....

Appendix D: Vocabulary Test

VOCABULARY TEST 1

Part One: Receptive Knowledge

Look at the 3 definitions. Choose the best word to match each definition and then write the word number next to the definition e.g.

1. cat		
2. skirt	***4***	occupation
3. bicycle	***2***	item of clothing
4. job	***1***	animal
5. paper		
6. star		

1. marble
2. palm ___ inner surface of your hand
3. ridge ___ excited feeling
4. scheme ___ plan
5. statue
6. thrill

1. apply
2. elect ___ choose by voting
3. jump ___ become like water
4. manufacture ___ make
5. melt
6. threaten

1. blame
2. hide ___ keep away from sight
3. hit ___have a bad effect on something
4. invite ___ ask
5. pour
6. spoil

1. bench
2. charity ___ part of a country
3. fort ___ help to the poor
4. jar ___ long seat
5. mirror
6. province

1. basket
2. crop ___ money paid regularly for doing a job
3. flesh ___ heat
4. salary ___ meat
5. temperature
6. thread

1. apparatus
2. compliment ___ set of instruments or machinery
3. revenue ___ money received by the government
4. scrap ___ expression of admiration
5. tile
6. ward

1. auspices
2. casualty ___ being away from other people
3. froth ___ someone killed or injured
4. haunch ___ noisy and happy celebration
5. revelry
6. seclusion

1. desolate
2. fragrant ___ good for your health
3. gloomy ___ sweet-smelling
4. profound ___ dark or sad
5. radical
6. wholesome

1. birth
2. dust ___ being born
3. operation ___ game
4. row ___ winning
5. sport
6. victory

1. auxiliary
2. candid ___ full of self-importance
3. dubious ___ helping, adding support
4. morose ___ bad-tempered
5. pompous
6. temporal

1. discharge
2. encounter ___ use pictures or examples to show the meaning
3. illustrate ___ meet
4. knit ___ throw up into the air
5. prevail
6. toss

1. annual
2. blank ___ happening once a year
3. brilliant ___ certain
4. concealed ___ wild
5. definite
6. savage

1. alcohol
2. apron ___ cloth worn in front to protect your clothes
3. lure ___ stage of development
4. mess ___ state of untidiness or dirtiness
5. phase
6. plank

1. circus
2. jungle ___ speech given by a priest in a church
3. nomination ___ seat without a back or arms
4. sermon ___ musical instrument
5. stool
6. trumpet

1. original
2. private ___ complete
3. royal ___ first
4. slow ___ not public
5. sorry
6. total

1. bruise
2. exile ___ agreement using property as security for a debt
3. ledge ___ narrow shelf
4. mortgage ___ dark place on your body caused by hitting
5. shovel
6. switch

1. blend
2. devise ___ hold tightly in your arms
3. embroider ___ plan or invent
4. hug ___ mix
5. imply
6. paste

1. blaspheme
2. endorse ___ give care and food to
3. nurture ___ speak badly about God
4. overhaul ___ slip or slide
5. skid
6. straggle

1. acquiesce
2. contaminate ___ work at something without serious intentions
3. crease ___ accept without protest
4. dabble ___ make a fold on cloth or paper
5. rape
6. squint

1. accident
2. choice ___ having a high opinion of yourself
3. debt ___ something you must pay
4. fortune ___ loud, deep sound
5. pride
6. roar

1. coach
2. darling ___ a thin, flat piece cut from something
3. echo ___ person who is loved very much
4. interior ___ sound reflected back to you
5. opera
6. slice

1. anterior
2. concave ___ small and weak
3. interminable ___ easily changing
4. puny ___ endless
5. volatile
6. wicker

1. dregs
2. flurry ___ worst and most useless parts of anything
3. hostage ___ natural liquid present in the mouth
4. jumble ___ confused mixture
5. saliva
6. truce

1. administration
2. angel ___ managing business and affairs
3. front ___ spirit who serves God
4. herd ___ group of animals
5. mate
6. pond

VOCABULARY TEST 1

Part Two: Productive Knowledge

Read the sentences and complete the underlined words e.g.

He was riding a bi<u>cycle.</u>

1. I’m glad we had this opp_____ to talk.

2. He has a successful car_____ as a lawyer.

3. The thieves have forced an ent_____ to the building.

4. Ann intro_____ her boyfriend to her mother.

5. Sudden noises at night sca_____ me a lot.

6. Her favourite flowers were or_____.

7. Many people are inj_____ in road accidents every year.

8. He wasn’t very popu_____ when he was a teenager, but he has many friends now.

9. We decided to celebrate New Year’s E_____ together.

10. Every working person must pay income t_____.

11. The evacu_____ of the building saved many lives.

12. You must wear a bathing suit on a public beach. You're not allowed to walk na_____.

13. He takes cr_____ and sugar in his coffee.

14. The government tried to protect the country's industry by reducing the imp_____ of cheap goods.

15. This is a complex problem which is difficult to compr_____.

16. In order to be accepted into the university, he had to impr_____ his grades.

17. She was sitting on a balcony and bas_____ in the sun.

18. The farmer sells the eggs that his he_____ lay.

19. The thieves threw ac_____ in his face and made him blind.

20. The dress you're wearing is lov_____.

21. They keep their valuables in a vau_____ at the bank.

22. We do not have adeq_____ information to make a decision.

23. Computers have made typewriters old-fashioned and obs_____.

24. She showed off her sle_____ figure in a long narrow dress.

25. La_____ of rain led to a shortage of water in the city.

26. The prisoner was released on par_____.

27. She found herself in a pred_____ without any hope for a solution.

28. The dead bodies were mutil_____ beyond recognition.

29. She is not a child, but a mat_____ woman. She can make her own decisions.

30. The pirates buried the trea_____ on a desert island.

31. I wouldn't hire him. He is unmotivated and indo_____.

32. Soldiers usually swear an oa_____ of loyalty to their country.

33. A bird perched at the window led_____.

34. To improve the country's economy, the government decided on economic ref_____.

35. Some coal was still smoul_____ among the ashes.

36. Don't pay attention to this rude remark. Just ig_____ it.

37. The lawyer gave some wise coun_____ to his client.

38. Teenagers often adm_____ and worship pop singers.

39. The insect causes damage to the plants by its toxic sec_____.

40. The voter placed the ball_____ in the box.

Appendix E: Questionnaire

PART 1: YOUR CONTACT WITH ENGLISH OUTSIDE THE CLASSROOM

	Please tick (✓) one box for each sentence.	**This is very untrue of me**	**This is not really true of me**	**This is neither true nor untrue of me**	**This is quite true of me**	**This is very true of me**
1.	I read English newspapers/magazines outside of class.	❑	❑	❑	❑	❑
2.	In England, outside of class, I speak English to my teacher/instructor.	❑	❑	❑	❑	❑
3.	I speak in English to service personnel (e.g. shop assistant, bank clerk).	❑	❑	❑	❑	❑
	3a. Please give more information *e.g. In what places do you most commonly speak to service personnel? (Please give names of places). What do you speak to service personnel about? (Please give examples).*					
4.	I use English outside of class to clarify classroom-related work.	❑	❑	❑	❑	❑
5.	I read e-mails or Internet web pages in English outside of class.	❑	❑	❑	❑	❑
6.	I use English for short exchanges with the people I live with (e.g. greetings / “Please pass the salt” / “I’m leaving”).	❑	❑	❑	❑	❑
7.	I listen to other people’s conversations in English outside of class.	❑	❑	❑	❑	❑
8.	I listen to English TV/spoken radio programmes/podcasts outside of class.	❑	❑	❑	❑	❑

	8a. Please give more information *e.g. Where are you when you listen to English TV/radio/podcasts? When do you usually do this? Who are you with? What kind of programmes/podcasts do you listen to? Do you do anything while listening? What?*					
9.	I speak *my native language* to native or fluent speakers of *my native language.*	❑	❑	❑	❑	❑
10.	I read novels in English outside of class.	❑	❑	❑	❑	❑
11.	I speak English outside of class with native or fluent English speakers who I do not live with.	❑	❑	❑	❑	❑
	11a. Please give more information *e.g. Who are the native/fluent English speakers you speak to? Where do you speak to them? When do you speak to them? What do you speak about? What are the speakers doing?*					
		This is very untrue of me	**This is not really true of me**	**This is neither true nor untrue of me**	**This is quite true of me**	**This is very true of me**
12.	I speak *my native language* to non-native speakers of *my native language.*	❑	❑	❑	❑	❑
13.	I write emails in English outside of class.	❑	❑	❑	❑	❑
14.	Overall, I think I read a lot in English outside of class.	❑	❑	❑	❑	❑
15.	I have long conversations in English.	❑	❑	❑	❑	❑
16.	I do long pieces of writing (e.g. more than one paragraph) for homework in English outside of class.	❑	❑	❑	❑	❑

	16a. Please give more information *e.g. Where do you usually do your writing? When? What objects do you use? Who are you with? Do you do anything while you write? What?*					
17.	Overall, I think I listen to a lot of English outside of class.	❑	❑	❑	❑	❑
18.	I read subtitles while I am watching TV or movies in English outside of class.	❑	❑	❑	❑	❑
19.	I use English to get directions or information from strangers (e.g. “Where is the post office?” / “How much are stamps?”).	❑	❑	❑	❑	❑
	19a. Please give more information *e.g. In what places do you most commonly get information from strangers? (Please give names of places). When does this usually happen? What kind of information do you get? (Please give examples).*					
20.	I read timetables, announcements, posters, menus etc. in English outside of class.	❑	❑	❑	❑	❑
21.	I speak *English* to non-native speakers of *my native language.*	❑	❑	❑	❑	❑
22.	I listen to English movies and videos outside of class.	❑	❑	❑	❑	❑

	22a. Please give more information *e.g. Where are you when you listen to English movies/videos? When do you do this? Who are you with? Do you do anything while you listen? What?*					
		This is very untrue of me	**This is not really true of me**	**This is neither true nor untrue of me**	**This is quite true of me**	**This is very true of me**
23.	I write personal notes or letters in English outside of class.	❑	❑	❑	❑	❑
24.	I speak a language other than English or my native language to speakers of that language (e.g. Swedish with a Swedish-speaking friend).	❑	❑	❑	❑	❑
25.	Overall, I think I do a lot of writing in English outside of class.	❑	❑	❑	❑	❑
26.	I speak *English* to native or fluent speakers of *my native language.*	❑	❑	❑	❑	❑
27.	I listen to English songs outside of class.	❑	❑	❑	❑	❑
28.	I speak English to the people I live with.	❑	❑	❑	❑	❑
	28a. Please give more information *e.g. Where are you usually when you speak English to the people you live with? Which room/place? When do you do this? What do you speak about? Who is there? What are the speakers doing?*					

PART 2: WHAT YOU DO WITH NEW ENGLISH WORDS

	Please tick (✓) one box for each sentence. *Outside of class*, when I hear or see a new word/phrase in English, I…	This is very untrue of me	This is not really true of me	This is neither true nor untrue of me	This is quite true of me	This is very true of me
1.	repeat the word/phrase in my head to try to remember it	❑	❑	❑	❑	❑
2.	use a bilingual dictionary to find the translation of the word/phrase in my language	❑	❑	❑	❑	❑
3.	use an English-only dictionary to find the meaning of the word/phrase	❑	❑	❑	❑	❑
4.	ask the speaker to explain (in English) the meaning of the new word/phrase	❑	❑	❑	❑	❑
5.	try to guess the meaning	❑	❑	❑	❑	❑
6.	try to work out what part of speech the word/phrase is (e.g. noun/verb/adjective)	❑	❑	❑	❑	❑
7.	try to remember the word/phrase to ask my teacher/friends later for the meaning	❑	❑	❑	❑	❑
8.	try to imagine how the word/phrase is spelt	❑	❑	❑	❑	❑
9.	ask the speaker to translate the word/phrase into my native language	❑	❑	❑	❑	❑
10.	use an English-only dictionary to find the word/phrase used in a sentence	❑	❑	❑	❑	❑
11.	write down the word/phrase and its meaning *immediately*	❑	❑	❑	❑	❑
12.	try to think of a translation of the word/phrase in my native language	❑	❑	❑	❑	❑
13.	ask the speaker to give me a synonym of the word/phrase in English	❑	❑	❑	❑	❑
14.	write down the word/phrase and its meaning *later*	❑	❑	❑	❑	❑
15.	decide whether or not I want to remember/learn the word/phrase	❑	❑	❑	❑	❑

	Outside of class, when I hear or see a new word/phrase in English, I…	This is very untrue of me	This is not really true of me	This is neither true nor untrue of me	This is quite true of me	This is very true of me
16.	try to think of a picture in my head to help me remember the word/phrase's meaning	❑	❑	❑	❑	❑
17.	try to think of English words/phrases that have a similar/opposite meaning	❑	❑	❑	❑	❑
18.	try to use the new word/phrase in a sentence as soon as possible	❑	❑	❑	❑	❑
19.	try to connect the new word/phrase to a personal experience I have had	❑	❑	❑	❑	❑
20.	write down how the word/phrase was used (e.g. the whole sentence) *immediately*	❑	❑	❑	❑	❑
21.	set myself goals for learning/using the word/phrase	❑	❑	❑	❑	❑
22.	try to think of situations where I could use the new word/phrase	❑	❑	❑	❑	❑
23.	say the word/phrase aloud to help me remember it	❑	❑	❑	❑	❑
24.	draw a picture/diagram to help me the remember the word/phrase and its meaning	❑	❑	❑	❑	❑
25.	write down how the word/phrase was used (e.g. the whole sentence) *later*	❑	❑	❑	❑	❑
26.	think about what words could be used before and after the word/phrase in sentences	❑	❑	❑	❑	❑
27.	try to decide the best way to learn the word/phrase	❑	❑	❑	❑	❑
28.	use any available pictures/body language to help me understand the word/phrase	❑	❑	❑	❑	❑

PART 3: INFORMATION ABOUT YOU

1. What is your name? _____

2. What is your e-mail address? (Please write clearly)

_____

3. What is your mobile phone number in England? _____

4. Are you…? Male ❑ Female ❑

5. How old are you? 18–24 ❑ 25–31 ❑ 32–38 ❑ 39– 45 ❑ 46– 52 ❑ 51– 55 ❑ 56 + ❑

6. Where were you born? (Please write name of country) _____

7. What is your native language? _____

8. What language(s) do you speak at home in your country? _____

9. Before now, have you ever been to an English-speaking country *for the purpose of studying English*? Yes ❑ No ❑

10. In what language(s) did you receive most of your *primary* education? (E.g. what language did your teachers use in class most of the time?)

		Number of years
Language 1		
Language 2		
Language 3		

11. In what language(s) did you receive most of your *secondary* education? (E.g. what language did your teachers use in class most of the time?)

		Number of years
Language 1		
Language 2		
Language 3		

12. When did you arrive in England? _ _ / _ _ / _ _ _ _ (DD/MM/YYYY)

13. In total, how long are you going to stay in England? _____

14. On what date (if known) will you leave England? _ _ / _ _ / _ _ _ _ (DD/MM/YYYY)

15. Please give details about the English language course(s) you are taking during your stay in the UK:

Type of class (e.g. general English, academic writing)	**Level** (e.g. B2, pre-intermediate)	**Number of hours per week**

16. Please tick (✓) ONE of the sentences (a – f) that best describes where you live in England at the moment.

a. ***I'm living in the home of an English-speaking family.*** ❑
ii. Do they speak your native language? Yes ❑ No ❑

b. ***I'm living in a student dormitory.*** ❑
i. Do you have a private room? Yes ❑ No ❑
ii. Do you have a roommate who is a native or fluent English speaker? Yes ❑ No ❑
iii. Do you live with people who are NOT native or fluent English speakers? Yes ❑ No❑

c. ***I'm living alone in a room or an apartment/flat.*** ❑

d. ***I'm living in a room or an apartment/flat with native or fluent English speaker(s).*** ❑

e. ***I'm living in a room or an apartment/flat with others who are NOT native or fluent English speaker(s).*** ❑

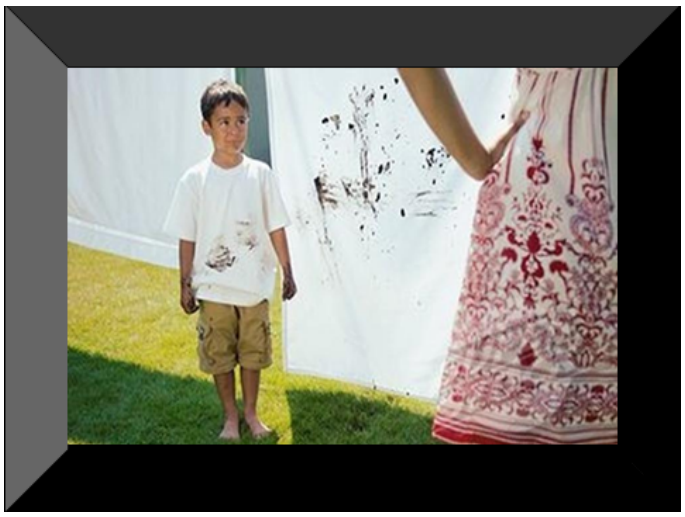
f. ***Other*** (please give details):

_____
_____
_____

Thank you very much for completing this questionnaire!

Appendix F: An example of an OWLS simulation

Cinematic music



Woman: [Gasp] What have you done? Oh, you're being such a nuisance today!

Appendix G: Interview Schedule

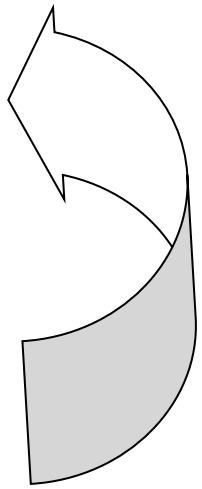
What would you do in this situation?

How exactly would you do that? Can you give me an example?

Why would you do that?

How do you think that would be helpful to you?

Would you do anything else? If yes, what?



Appendix H: Strategies taxonomy

Classification	Strategic Behaviour
Planning	• Choose to work with word/phrase or not • Choose to learn word/phrase or not • Choose which aspects of word/phrase to learn
Determination: Individual	• Analyse part of speech to determine the function of word/phrase • Analyse affixes and roots to determine part of speech • Check for L1 cognate/translation to transfer L1 knowledge • Check for parallels with languages other than the L1 to link with previous knowledge • Analyse any available pictures, gestures or sounds to enhance comprehension of meaning • Guess from looking at the word • Guess from textual context (immediate context) • Guess from background knowledge (wider context) • Use a bilingual dictionary to link to L1 knowledge • Use a monolingual dictionary to find meaning • Use a monolingual dictionary to find word/phrase in context • Consult other reference source (e.g. search engine) to find meaning • Consult other reference source to find word/phrase in context • Consult other reference source to find pictorial representation • Repeat new word/phrase in head to aid recognition
Determination: Seeking help of others	• Ask interlocutor for L1 translation (immediate) • Ask for repetition of new word/phrase to aid recognition • Ask interlocutor for paraphrase or synonym of new word/phrase (immediate) • Ask interlocutor for meaning of new word/phrase (immediate) • Ask L1 speaker present for L1 translation (immediate) • Ask L1 speaker present for L1 translation (delayed) • Ask native speaker present for meaning of new word/phrase (immediate) • Ask native speaker for meaning of new word/phrase (delayed) • Ask non-native speaker present for meaning of new word/phrase (immediate) • Ask non-native speaker for meaning of new word/phrase (delayed)
Memorisation	• Make a note of the usage of word/phrase (immediate) • Make a note of the usage of word/phrase (delayed) • Make a note of the meaning of word/phrase (immediate) • Make a note of the meaning of word/phrase (delayed) • Ask for repetition of new word after meaning has been determined • Connect word/phrase to personal experience

	• Create pictorial representation of word/phrase meaning • Image word/phrase meaning • Image word/phrase spelling • Associate word/phrase with its coordinates • Connect the word/phrase to its synonyms and antonyms • Say word/phrase aloud in isolation • Say word/phrase in a sentence • Repeat word/phrase silently 'in your head'
Metacognition	• Setting goals for learning word/phrase • Decide how best to learn word/phrase • Plan opportunities to encounter/use word/phrase • Setting goals for number of words/phrases learnt over a period of time • Evaluate success of vocabulary learning against goals

Appendix I: OWLS interview transcript pair

PARTICIPANT M1

Interviewer: This is Kxxx at the LSCE on the 11th of August 2013. Okay, so I'm going to show you the first one.

Participant: Okay.

Interviewer: Errr...slideshow? Here we go. Can you see?

Participant: Yes, can see.

Simulation 1 plays.

Participant: You need me to say something?

Interviewer: Well, um, did you hear what the man said?

Participant: Kind of. What do you feel or something.

Interviewer: What do you feel...?

Participant: What do you feel about the English school?

Interviewer: Nearly!

Participant: Can I to hear it again?

Interviewer: Yeah, sure.

Simulation 1 plays again.

Interviewer: What did you get that time?

Participant: What do you...fell? Fellow like? He said 'fellow'?

Interviewer: Fellow, yeah.

Participant: Okay. I need to answer this?

Interviewer: Would you? Well, what would you say?

Participant: Ah, okay, er, I, I think is like my partners or my classmate.

Interviewer: Erm, yeah. Would you ask him, "Do you mean my partners or my classmates?" or would you just answer the question?

Participant: Errr, because I don't, I'm not sure what is the meaning of 'fellor'.

Interviewer: Fellow.

Participant: And because of maybe I try to ask again like "What do you mean? My classmate?"

Interviewer: Okay.

Participant: Maybe. I try to understand what he is trying to say to me.

Interviewer: Okay.

Participant: Maybe I'm not sure about the fellows, I don't know is like classmates.

Interviewer: So you'd ask him if he means 'classmates'?

Participant: Yes.

Interviewer: And then would you do anything else?

Participant: And, er, maybe I, I after that I can answer something. Maybe they, they are very international, the most people are from Europe, erm, something like this. And Arabic or Chinese, Asian. I, and I try to answer this after, try to understand what he try to say to me.

Interviewer: That's fantastic. Thank you.

Interviewer: So, I could say, do you have any fellow Venezuelans at London School of English?

Participant: No.

Interviewer: No? You're the only one?

Participant: I'm the only one!

Interviewer: Oh! Representing!

Participant: [Laughs] Yes! In about six months I have, I have, I have been the only one.

Interviewer: Wow!

Participant: Yes.

Interviewer: Okay.

Participant: Too many people from Brazil.

Interviewer: Too many? Or a lot of?

Participant: Yes, a lot [laughs]. A lot of people from Brazil. And maybe from Colombia, when Axxxx and Axxxx, where, did you. But from Brazil. No, only me.

Interviewer: Oh, that's nice. Okay, brilliant, so let's move on to the next one.

Simulation 2 plays.

Participant: Ok, er, I, do you want to explain what I understand?

Interviewer: Yeah.

Participant: Er, in the movie, she said, er, like, to his son, 'what have you done'? Like, you, you are very dirty, and er, I, I'm not sure what, the, the word or the adjective that she used but I think

for the context she's like, er, what are you, have, what, 'what have you done?' Like, where have you been? Why are you like this, of dirty?

Interviewer: Okay, and what is it about the context which tells you that?

Participant: Er, because is, is, the kid is of, errr, black, I don't know, the hands, the clothes. I think the word of the, when something is, this, I don't know this word.

Interviewer: Erm, I suppose, like, handprints?

Participant: But when, if you're drink wine, and you pfft [mimes spilling wine on trousers].

Interviewer: Oh, spilt!

Participant: Ah, spilt!

Interviewer: To spill.

Participant: Yes. I think because he's very dirty, dirty, and her mother, his, his mother is er, is very mad.

Interviewer: Great! Okay, so you're in the cinema, you're watching this, you don't hear the word or you don't quite understand the word but you can guess from the context...

Participant: Yes, I, yes. I imagine.

Interviewer: So, would you in, you leave the cinema, would you do anything with this word or would you just forget about it, what?

Participant: Mmm, mmm, to be honest, I think I forget.

Interviewer: Okay.

Participant: Because I, I, it's no one close to me that can explain me, maybe I could, I go out and I think about everything, not in this word, and I can forget about this word.

Interviewer: Okay. Thank you very much. Right, number three.

Simulation 3 plays.

Participant: Ok, er, they are ask the people, asking the people, er, who wants to do the IELTS that they need to came to this class preparation. And it's very urgent to people who wants to do that.

Interviewer: Okay. And so this...? [Points to word in red]

Participant: Is urgent. Is urgent. Is...

Interviewer: Okay. So it needs to be done quickly?

Participant: Yes, quickly. Something quickly, to, to, to come quickly to this preparation class.

Interviewer: Ah, okay.

Participant: Oh, yes.

Interviewer: How do you know the meaning?

Participant: Ah, I know urgent. I know this word.

Interviewer: Okay, excellent. Er, number four?

Simulation four plays.

Participant: She asked him about this product is on sale, and he said that you need to go to the section of, of sale, selling, of sales, bargain.

Interviewer: Bargain.

Participant: Bargain.

Interviewer: Yeah.

Participant: Okay, I know. So, I go to this section and look for this thing.

Interviewer: Would you do anything else?

Participant: Ah, no. No, I think.

Interviewer: Okay, good stuff. Here's number five.

Simulation 5 plays.

Participant: Ah, I don't understand what he said like!

Interviewer: Do you know the song?

Participant: Yes, but I don't know who is the singer.

Interviewer: It's Carly Simon. I think it's from like 1960s. It's old!

Participant: Ah!

Interviewer: Would you like to hear that bit again?

Participant: Yes.

Interviewer: Okay.

Participant: I have heard this, but...

Simulation 5 plays again.

Participant: You're so vain? This song is about you? This song is about you. You are so vain? Vain?

Interviewer: You're so vain.

Participant: You're so beautiful?

Interviewer: Ah, no!

Participant: No? Vain, what is?

Interviewer: What do you think it might mean?

Participant: Ah! Yes, I know this word. It's the same as Spanish. Yes, yes, I know this word. But is not good!

Interviewer: No! No. So, this, I think Carly Simon wrote this song about an ex-boyfriend, um [Laughs].

Participant: [Laughs] Okay.

Interviewer: So when you listen to music in English, um, so say for example you are this girl and you're listening to this song, would you just enjoy the song, or would you try to find the word that you weren't sure of, or...

Participant: If the sound I heard I like, I try to, to find the meaning of the word, or, but if is something is common or something that I not interesting, interested, I, I, yes, I don't care, but if I like, if I like I try to find the correct meaning and the correct words.

Interviewer: How do you do that?

Participant: Ah, in the Internet.

Interviewer: Okay. So you go to what kind of website?

Participant: Maybe, I don't know, yeah, I tend to put the name of the song.

Interviewer: Okay.

Participant: Um, put the, I don't know, letter, or, I don't know, something like this.

Interviewer: Okay, nice. Thank you very much. And three more. Here's number six.

Simulation 6 plays.

Participant: Okay, she was looking, she was watching the news. And, err, it was the, the queen arrived in Manchester hospital, Manchester's, and err, many people turns out, turns out? To receive her, and err, okay, that's it.

Interviewer: Okay, perfect! And 'turned out' means what?

Participant: Um, it turns it, errr, people arrive, they were in the place, appear.

Interviewer: Okay. And have you heard that phrasal verb before?

Participant: Yes, yes. I have a doubt that today, but I forget to ask! I don't like when a, some person it said to you in a rude way, like, you need to do this, you [*Wags finger*]...what is? Put off? Pull out? I don't know the phrasal verb. I forget.

Interviewer: Um, if I say, 'Oh, you're a very bad student – you must do your homework, you must try harder'?

Participant: Uh-huh, uh-huh.

Interviewer: To tell off.

Participant: Tell off! Ohh, yes!

Interviewer: To reprimand. To tell off.

Participant: Okay, thank you!

Interviewer: Okay. Good phrasal verb!

Participant: Yes, I was trying to remember but, oh ((I, er))...

Interviewer: Phrasal verbs are so hard.

Participant: Yes! Everything is off, off, up, down [*Laughs*].

Interviewer: [*Laughs*] Yep, yep. Okay.

Participant: Okay.

Interviewer: Well done with that one.

Participant: Thank you.

Interviewer: Right, and the penultimate.

Simulation 7 plays.

Participant: Okay she was cooking, and, the, her phone rang, and she ask someone to, to take the phone to her while she was answer.

Interviewer: Okay. Do you want to listen again, see if you still agree?

Participant: Okay.

Simulation 7 plays again.

Participant: She ask, 'Can you take over here while I answer it'.

Interviewer: Mmmhmm.

Participant: So, that mean that she, she didn't answer, she. No.

Interviewer: She said can you, can you, so she said 'Ooh that's my phone, can you take over here so I can answer it, while I answer it'.

Participant: So, I, I, you ask me and I give you the phone?

Interviewer: Ah, okay – so that would be 'bring'. Can you bring me the phone?

Participant: Ahh.

Interviewer: But here she says, 'Can you take over, here, while I answer it'.

Participant: And she want me to answer it.

Interviewer: No – 'to take over' do you understand the meaning of it?

Participant: Take over, I, I think that is to take, is something like you bring.

Interviewer: Okay. And the final one. Here we go. Ahh – sorry.

Simulation 8 plays.

Participant: The advertisement means that, pick, pick up, is choose, like, choose. Choose a Yum bar today, to buy, buy this Yum bar, delicious! [*Laughs*].

Interviewer: [*Laughs*] Buy this delicious Yum bar!

Participant: Okay.

Interviewer: And, do you know the phrasal verb pick up, or was that just a guess?

Participant: I, I, I, I saw this today.

Interviewer: Oh yeah? Can you remember where?

Participant: Yes, he said like, errr, when I was young, my teacher used to pick up me, pick up on me, no.

Interviewer: And what did it mean?

Participant: No, no it was different – sorry!

Interviewer: What did the teacher do?

Participant: No, no, it was pick on, is not pick up.

Interviewer: Ah, pick on.

Participant: Pick on me, err, pick up, er, try, maybe.

Interviewer: Try?

Participant: Try a Yum bar today?

Interviewer: Yeah, it's a good guess.

Participant: Um, get?

Interviewer: Yeah.

Participant: Ah, okay. Okay.

Interviewer: Alright. Thank you so much.

Participant: Thank you.

PARTICIPANT M2

Interviewer: This is the interview with Mxxx on the 6th December at HFC. Are you ready? Here's the first one.

Simulation 1 plays.

Participant: Huh?

Interviewer: What's happening in here? [*Points to head*]

Participant: Eh? What should I do?

Interviewer: Well, um, you're in, imagine you're in this situation and um, this, um, young man turns to you and says this. What do you do? What do you say, or what are you thinking?

Participant: Mmmm, I think, mm, he wants to talk about, about, lives, or, yesterday things, or tomorrow friend...

Interviewer: Yeah? Why do you think that?

Participant: Mmm, um, mm, everyday, when I eat, where I have lunch or dinner, I always talk about, um, friends, so I think, yeah.

Interviewer: Ok, great. So these people are? Who are these people?

Participant: [*Laughs*] Errrr...students, yeah.

Interviewer: Students. Okay. And did you hear what he said?

Participant: Er, I, I couldn't hear. Listen again, please?

Interviewer: Sure

Simulation 1 is repeated.

Participant: Mmm, language? Er, what's, what's do you speak language?

Interviewer: Okay. So, imagine that you're in this situation, imagine that you're sitting here, next to this lady [*Points to screen*], and the man turns to you and he says this, and you only hear 'language' or you're not sure, what would you do?

Participant: Mmmm, err, I ask, I ask, could, can you repeat that? Mmm.

Interviewer: Okay. You ask him to repeat the question and you hope that you can hear it again.

Participant: Yes.

Interviewer: Okay. Would you do anything else?

Participant: Mmmm. No.

Interviewer: No? Okay, good stuff. Alright, here's number two.

Simulation 2 plays.

Participant: Hmm, the boy played on the ground so, so his clothes became dirty so his mother, a little angry [*Laughs*].

Interviewer: Uh-huh. And how do you know all of this? How do you know she's angry?

Participant: Er, I, I heard her voice and this arm [*Mimes hand on hip and laughs*]

Interviewer: [*Laughs*] Oh, she's got her hand on her hip, okay, yes! The angry pose! And did you hear what she said?

Participant: Mmm, what's, what have you done, and I couldn't hear.

Interviewer: Okay, don't worry. And where is this situation?

Participant: Mmm, ah, mmm, his or family's garden.

Interviewer: Oh okay, yes, yes, cause you've got the grass and the washing hanging up and stuff. What about the people in slide 2?

Participant: In the cinema?

Interviewer: Uh-huh.

Participant: Ah! On the movie.

Interviewer: Yes. So imagine that you are one of these people sitting in the cinema, eating your popcorn, drinking your Coca-Cola, and you're watching this movie and you see this scene, this part of the movie, and you didn't under-, or you didn't hear everything that the mother said, for you is this a problem or do you feel comfortable that you understand the scene?

Participant: Mmm, if I couldn't hear all sentence but I, I can watch the screen so I can guess the, er, guess the situation, so I think I, I'm, I don't have to hear all sentence.

Interviewer: Okay. Do you go to the cinema much in Oxford?

Participant: No.

Interviewer: Have you ever been to the cinema in Oxford?

Participant: Um, no, I have never been to cinema in Oxford.

Interviewer: Okay, and is there a particular reason for that?

Participant: Um, no, but I want, I would like to go cinema but, I, I have no time to go, so.

Interviewer: Ah! Too busy giving presentations, writing essays?

Participant: Yeah!

Interviewer: Okay, well done. Here comes number three.

Simulation 3 plays.

Participant: Hmm. Er, I think, mm, the, the students were, students are watching the board, ah, and, er, they talk about, they talk about this things, so, and, er some people think about, er whether I, I go, I got IELTS exam, mm, and other people think about they go to presentation on Wednesday so maybe, mm, some people, invite, some people invite, invite? Invite someone.

Interviewer: Okay. And do you understand all of the language on the poster?

Participant: Yes.

Interviewer: Yeah?

Participant: Yeah.

Interviewer: What does the word in red mean?

Participant: Ummm, 'urgey' is, mm, I think we have to, have to do something more emphasize, so near must doing, do, must do, I think.

Interviewer: How do you know that that's the meaning?

Participant: Mmm, the word, word is red, mmm, more important, yes, so.

Interviewer: Okay, thank you. And if you were a student of this school and you were going to take, um, the IELTS exam at the end of this month, what would you do?

Participant: Erm, I, I study, I will study, and, yes. If I, if it is first time to take IELTS exam, I will search the IELTS tests things.

Interviewer: Okay, thanks. Here's the next one.

Simulation 4 plays.

Participant: Mmm, mm, she, she want to buy the beans, mmm, but, mmm, she want to discount the beans and so she ask the customer to discount the beans, yes.

Interviewer: Okay, and what did he say?

Participant: Mmm, mm, he, he want to discount but, don't, she, he can't discount.

Interviewer: Okay. Yes, you were right – she wanted a discount – and what did he say?

Participant: Bargain section.

Interviewer: Do you now the word bargain?

Participant: Bargain is a good shopping?

Interviewer: Er, yes, yeah. So what would you find in the bargain section of a supermarket, do you think?

Participant: Section? Bargain corner? Only bargain of shop?

Interviewer: Yeah. And you said bargain is 'good shopping' – what makes a bargain a good buy?

Participant: Mmm, er, mmm, not sure.

Interviewer: Okay, not to worry. Next one.

Simulation 5 plays.

Participant: Mmm, she is listening to music, mm, but I don't understand why the song change [*Laughs*].

Interviewer: Oh, it's the same song!

Participant: Yes?

Interviewer: Yes, it's the same song, but it's different sections of the song.

Participant: Ah! Mmm.

Interviewer: Do you know the song?

Participant: No, I haven't, I have never seen, I have never listened.

Interviewer: Heard.

Participant: Ah, heard. Yeah.

Interviewer: It's quite an old song. I think it's from, like, the 1970s. An American song. And do you ever listen to English music? Or English language music?

Participant: Sometimes.

Interviewer: Yeah?

Participant: Mmm.

Interviewer: And what do you do in terms of the lyrics, the words, when you're listening?

Participant: Mmm. I want, I want to hear again.

Interviewer: This one again? Yes, of course. Yeah.

Simulation 5 is repeated.

Participant: Mmm, don't you?

Interviewer: Don't you, yeah.

Participant: Mmm, range?

Interviewer: Ah! It rhymes with 'range'.

Participant: Mmm. Again?

Simulation 5 is repeated.

Participant: Mmmm...Range?

Interviewer: Are you sure?

Participant: No.

Interviewer: You said sometimes you listen to English music: can you imagine that you're listening to music and you hear this song, and you hear 'don't you' and you hear 'range' or something that sounds like range, and you want to find this song because you like it, you want to download it, what would you do with those words that you heard? How would you find the words, or how would you understand the meaning of the words?

Participant: Uhhh, mmm, er, I will search on the internet.

Interviewer: Where?

Participant: Mmm, Google. If I know the title, I, I, I search the title on the Google. Uh-huh. Or if I know the singer, mm, I search, the singer's information in the internet. Yeah.

Interviewer: And if you don't know the title or the singer?

Participant: Mmm. Errr, mmm, I, I will, I will listen many times and, er, mmm, so, gradually I, I can understand the words and words, so, mm, but if I, if I can't understand the meaning of them, so I search the learn word, in the words, in the internet.

Interviewer: Okay. Good idea. Here's number six.

Simulation 6 plays.

Participant: She watch, she was watching BB, BB news, and this announcer said about the queen visitation. Erm, queen, erm, visited new children's hospital, so, so many children welcomed her.

Interviewer: Okay, great. Great. Is there anything that you didn't understand in what the newsreader said?

Participant: Hmmm, I want to listen again.

Interviewer: Yeah? Okay.

Simulation 6 is repeated.

Participant: Mmmm, turned, turned on, turn, turn on?

Interviewer: Not turn on...

Participant: Turn out? Means, er, mm, hmm, I don't understand.

Interviewer: Well, erm, what was the full sentence? What sentence did it come in?

Participant: Mmm, children turned out the street.

Interviewer: Yeah...

Participant: Mmm. Mmm. Turn. Mmmm, I think many people mmm, made a, made a queue, er, so queen walked next to the queue.

Interviewer: Yeah? So there's many people on either side of the street?. And how did you know that?

Participant: Eh, mmm, I, I see, I think about the situation, if I, if queen visits somewhere. And in Japan also like many people made a queue, so, I, I thought this.

Interviewer: Okay. Good reasoning.

Participant: Ah, okay.

Interviewer: Okay, good stuff. Here's number seven.

Simulation 7 plays.

Participant: Mmm, while she cuttings, cutting vegetables, um, the phone calling but she, she cannot take the call, mmm, mm, I, I think maybe they're, they are some people so she want to take, take phone somebody, person.

Interviewer: Okay, so imagine it's you, imagine she's talking to you, here she is, she's doing the chopping, you're there, and then she says this. How do you respond?

Participant: Yes, of course. Yeah.

Interviewer: Okay.

Participant: Could you, mm, can you repeat?

Interviewer: Of course.

Simulation 7 is repeated

Participant: Mmm, I, I answer, please means, ah! Er she, she will take the phone you, you have to, you have to have to take phone.

Interviewer: Okay, okay, good. And finally.

Simulation 8 plays.

Participant: Mmm, when I drive a car, I, I found this poster. The, this poster said welcome. This poster said Yum bar want to come, want to come some land, mm, so this is advertise.

Interviewer: Yep. Good. And do you understand all of the language on the advertisement?

Participant: Mmm, pick, pick up means, uh, select. Select or choose.

Interviewer: Have you seen this phrasal verb before?

Participant: Mmm, yes, mm.

Interviewer: Okay. So choose a Yum bar today?

Participant: Yes.

Interviewer: Okay, good. That's great. You did fantastic. Thank you so much.

Appendix J: Member check

2013/8/28 Jess Briggs <jess.briggs@kellogg.ox.ac.uk>

Dear Mxxx,

How are things with you?

As we discussed last week when we met, I have typed up our interview on a Word document and I have attached the document to this email.

I'd be very grateful if you could take some time to please read the interview document carefully and then reply to me by email, answering these three questions:

1. In your opinion, is this a truthful representation of our interview?
2. In your opinion, is this an accurate representation of our interview?
3. In your opinion, is this a comprehensive representation of our interview (i.e. is there anything that we said which I haven't included in the document?)

Thanks again for agreeing to help!

Yours sincerely and with very best wishes,

Jess

<G1_Transcript.docx>

Jessica G. Briggs
DPhil Candidate in Education
Teaching Assistant: MSc / PGDip in Teaching English Language in University Settings
http://www.education.ox.ac.uk/courses/pgdip-telus/
Department of Education
University of Oxford
jess.briggs@education.ox.ac.uk

From: Mxxxx Fxxxxx Bxxxxxxx [mxxxxfxxxxxx.bxxxxxx@gmail.com]

Sent: 02 September 2013 10:52

To: Jess Briggs

Subject: Re: Hello and the next stage of the research study

Dear Jess,

Hope you are great!! Sorry for the delay. I'm finishing the course on the 18th of September, so I can read it the next day of the week.

King Regards

<jess.briggs@kellogg.ox.ac.uk> wrote:

Hi Mxxx,

Many thanks for your reply.

OK, that's no problem – I look forward to hearing from you on the 18th September.

Good luck for end of the course!

Thanks again and all best wishes,

Jess

Jessica G. Briggs

DPhil Candidate in Education
Teaching Assistant: MSc / PGDip in Teaching English Language in University Settings
http://www.education.ox.ac.uk/courses/pgdip-telus/
Department of Education
University of Oxford
jess.briggs@education.ox.ac.uk

From: Mxxxx Fxxxxxxx Bxxxxxxx [mxxxxfxxxxxxx.bxxxxxxx@gmail.com]

Sent: 18 September 2013 12:12

To: Jess Briggs

Subject: Re: Hello and the next stage of the research study

Dear Jess,

I read the interview and now I feel embarrassed!! My English is very bad!!! For me the answer is yes for all the questions.

Regards

Sent from my iPhone

Appendix K: Regression analyses with mean informal contact variable

Hierarchical multiple regression analysis was employed to determine the ability of the singular informal L2 contact variable, which was entered as block 1, and length of stay (represented by the two dummy variables) and location of SAE, entered together as block 2, to predict the dependent variable of vocabulary gain. These analyses took receptive, productive and overall vocabulary gain as the dependent variable respectively.

K.1 Sample groupings

The continuous variable from the informal L2 contact measure, which comprised the means of scores in response to the 28 informal contact scenarios in part 1 of the questionnaire, was taken as the independent variable of primary interest. In this variable participants could score a minimum of 1 (if they had allocated '(1) This is very true of me' to every scenario) and a maximum of 5 (if they had allocated '(5) This is not at all true of me' to every scenario). In terms of the other two independent variables, the sample was divided into a three-level ordinal variable based on length of stay (group 1, 6-10 weeks; group 2, 11-15 weeks; group 3, 16-20 weeks) and into a binary categorical variable based on the location of the SAE (Oxford or London).

K.2 Parametric assumptions

Here follows a description of how the variables were investigated and treated to ensure they met the parametric assumptions required for regression analysis.

K.2a *Ratio of cases to independent variables*

To decide upon the minimum number of cases necessary to conduct multiple regression analysis Tabachnick and Fidell (2001) proffer the following formula, where *m* signifies the number of independent variables: $N>50+8m$. The receptive gain variable comprised 196 cases and the logarithmically transformed productive gain variable 129 cases. Due to participant mortality between time 1 and 2 of the vocabulary test, the overall vocabulary gain variable comprised 193 cases. All three dependent variables therefore exceed the ratio advised by Tabachnick and Fidell (2001).

K.2b *Treatment of outliers*

A discussion of the treatment of outliers detected in the overall and productive vocabulary gain variables is given in 4.1.1.4a and 4.1.1.4b respectively. An account of the logarithmic transformation of the productive vocabulary gain variable in response to outliers is provided in 4.1.1.4c.

One outlaying case was detected in the identification with informal L2 contact variable. This case was removed. No outlaying cases existed in either of the other two independent variables.

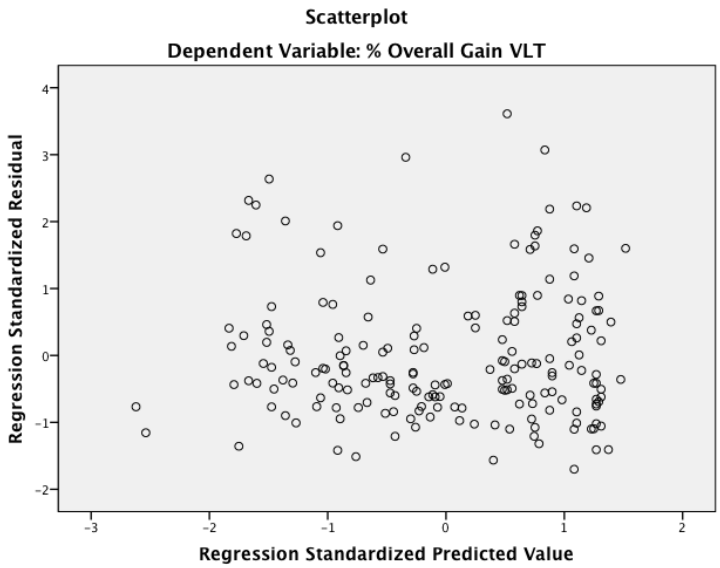
K.2c *Multicollinearity and singularity*

Correlation analyses were conducted on the three independent variables to be used in the regression model to find one significant correlation: there was a positive relationship between length of stay and identification with informal L2 contact [r=.246, N=240, *p*=.000]. Berry (1993) states that r values above .9 indicate that independent variables are too highly correlated for use in a regression model, and Tabachnick and Fidell (2001) suggest that r values as low as .5 may be problematic. Based on these figures in relation to the far lower r value found between two of the independent variables in this study, both variables were retained in the regression model.

K.2d *Normality, linearity and homoscedasticity of residuals*

An initial regression (taking overall vocabulary gain as the dependent variable) was run to check the assumptions of normality, linearity and homoscedasticity. Examination of the residuals scatterplot for the final model (figure K.1) in the hierarchical regression indicated that there was very little deviation from normality and that non-linearity and heteroscedasticity of residuals was not an issue.

Figure K.1: Scatterplot of regression residuals



K.2e *Independence of errors*

The Durbin-Watson statistic of the final model of the initial regression was 2.07, suggesting that the residuals were uncorrelated and therefore that the assumption of independence of errors had not been violated.

K.2f *Solution outliers*

In order to inspect the model for solution outliers, the Mahalanobis distances produced by the initial regression were investigated. Tabachnick and Fidell (2001) state that the critical

chi-square value for a model with four independent variables is 18.47 and that any cases that exceed this value can be deemed as solution outliers. No cases in the model had a Mahalanobis value exceeding 18.47.

K.3 Receptive vocabulary gain

A hierarchical multiple regression analysis was conducted to determine the extent to which the independent variables (block 1= identification with informal L2 contact scenarios; block 2= length of stay and location of SAE) were able to predict the receptive vocabulary gain scores. After the entry of block 1, the proportion of variance explained was 0%. After block 2 had been entered into the model, the model as a whole explained 3.1% of the variance. Neither model 1 [F(1, 193)=.035, *p*=.851] nor model 2 [F(4, 190)=1.508, p=.201] reached statistical significance.

Length of stay was the only predictor variable to make a statistically significant contribution to the regression model (β=-2.32, *p*=.021): this finding mirrors that of the ANOVA of these variables (reported in Appendix L) in which length of stay came close to reaching significance [F(2, 191)=2.86, *p*=.06] with a small effect size (Eta squared=.03). The negative beta value of this finding in the regression analysis indicates that membership of the medium stay group (11-15 weeks) in fact predicts lower vocabulary gain.

Table K.1: Predictor variable contribution to receptive vocabulary gain regression equation

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	17.973	13.261		1.355	.177
Questionnaire part 1 summed score	-.035	.186	-.014	-.188	.851
2 (Constant)	21.077	14.551		1.449	.149
Questionnaire part 1 summed score	-.028	.192	-.011	-.145	.885
Location of SAE	-.429	5.005	-.006	-.086	.932
Length of stay dummy variable 1: 6-10 wks?	-.333	6.388	-.004	-.052	.958
Length of stay dummy variable 2: 11-15 wks?	-14.356	6.187	-.175	-2.320	.021

K.4 Productive vocabulary gain

A hierarchical regression analysis was conducted to determine the extent to which the primary and other two independent variables were able to predict the logarithmically transformed productive vocabulary gain variable. After the informal L2 contact variable was entered, the overall model explained 1.8% of the variance in productive vocabulary gain, but this finding was not statistically significant [F(1, 127)=2.37, *p*=.13]. After block

2 had been entered (location and length of stay) the model as a whole explained 4.7% of the variance but again, this model did not reach statistical significance [F(4, 124)=1.51, *p*=.20].

Although none of the predictor variables made a statistically significant contribution to the final model, the predictor variable of primary interest, identification with L2 contact scenarios, came close (*β*=-.173, *p*=.059). As the beta-weight in this case is a negative value and lower scores in response to this variable signified higher identification (because 1=This is very true of me and 5=This is not at all true of me), this finding appears to suggest that those who identified more highly with informal contact scenarios gained more vocabulary. However, as this finding failed to reach significance it is merely an indication of a possible relationship in the data. In support of this indication, the ANOVA of these variables (reported in Appendix L) did find a significant effect for identification with informal L2 contact scenarios [F(1, 124)=5.76, *p*=.018], with a small effect size (Eta squared=.04).

K.5 Overall vocabulary gain

Hierarchical multiple regression was also conducted to determine the extent to which the four independent variables were able to predict the overall vocabulary gain scores. After block 1 had been entered the model explained 0% of the variance. After block 2 had been entered the model explained 3.8% of the variance. Neither model 1 [F=(1, 190)=.001, *p*=.97] nor model 2 [F(4, 187)=1.861, *p*=.12] attained significance. As in the regression analysis run with receptive gain as the dependent variable, the sole predictor variable to make a significant contribution in this model was length of stay (*β*=-2.65, *p*=.009): once more, belonging to the medium length of stay group (11-15 weeks) appeared to predict lower overall vocabulary gain. Again, the corresponding analysis of variance (Appendix L) found a supporting significant main effect for length of stay [F(2, 188)=3.62, *p*=.029] with a small effect size (Eta squared=.04). Post-hoc comparisons using the Scheffe test indicated that the mean vocabulary gain of the medium stay group (11-15 weeks; N=47, M=5.22, SD=21.85) was significantly different from the mean gain of the long stay group (16-20 weeks; N=101, M=18.04, SD=27.73).

Table K.2: Predictor variable contribution to overall vocabulary gain regression equation

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	14.178	10.387		1.365	1.74
Questionnaire part 1 summed score	-.006	.146	-.003	-.039	.969

2 (Constant)	21.084	11.354		1.857	.065
Questionnaire part 1 summed score	-.038	.150	-.019	-.255	.799
Location of SAE	-1.157	3.905	-.021	-.296	.767
Length of stay dummy variable 1: 6-10 wks?	-4.934	4.984	-.077	-.990	.323
Length of stay dummy variable 2: 11-15 wks?	-12.788	4.828	-.201	-2.649	.009

K.6 Summary of multiple regression analyses

The findings of the regression analyses here discussed indicate that the model as a whole did not significantly predict vocabulary gain. Only one predictor variable in the model, length of stay, made a statistically significant contribution: belonging to the medium stay group predicted lower receptive and overall vocabulary gain but had no significant relationship with productive vocabulary gain. There is evidence to suggest a small effect of identification with informal L2 contact scenarios on productive vocabulary gain.

Appendix L: ANOVA analyses of informal contact, length of stay and location on vocabulary gain

A series of three-way ANOVAs was conducted to explore the impact of each of the independent variables on receptive, productive and overall vocabulary gain respectively. The primary independent variable, identification with informal L2 contact scenarios, was represented as binary categorical variable dichotomised at the median split. Length of stay was a 3-level ordinal variable (short, medium or long), and location of SAE a dichotomous variable (Oxford or London).

L.1 Receptive vocabulary gain

The first ANOVA of the series took receptive vocabulary gain as the dependent variable to find no statistically significant effect for any of the independent variables. Length of stay came close to reaching significance [$F(2, 191)=2.86$, $p=.06$] with a small effect size (Eta squared=.03).

L.2 Productive vocabulary gain

The ANOVA conducted with the logarithmically transformed productive vocabulary gain variable as the dependent variable found a statistically significant main effect for identification with L2 contact scenarios [$F(1, 124)=5.76$, $p=.018$] with a small effect size (Eta squared=.04).

L.3 Overall vocabulary gain

The ANOVA that took overall gain as the dependent variable found a statistically significant main effect for length of stay [$F(2, 188)=3.62$, $p=.029$], however the effect size was small (Eta squared=.04). Post-hoc comparisons using the Scheffe test indicated that the mean vocabulary gain of the medium stay group (11-15 weeks; N=47, M=5.22, SD=21.85) was significantly different from the mean gain of the long stay group (16-20 weeks; N=101, M=18.04, SD=27.73). The short stay group did not differ significantly from either other group. The main effects for identification with L2 contact scenarios and location of SAE did not reach statistical significance.

Appendix M: Regression analyses with mean strategic behaviour variable

A regression model was run in which the five strategic behaviour factor variables were replaced in block 1 by the original strategic behaviour variable (which was made up of the mean scores in response to the items in part 2 of the questionnaire). To mirror the analyses described in 4.3.2, the strategic behaviour variable was also dichotomised for use in a series of corresponding ANOVAs in order to investigate whether ANOVA would support the findings of the regression analyses. For the ANOVAs, the strategic behaviour variable was transformed to make a binary nominal variable using a median split based on the sum of the mean of participants' reported strategic behaviour from the questionnaire; those with lower means rated the most number of statements as being very true of them and were thus labelled the High Identifiers (HI), and those with higher means were labelled the Low Identifiers (LI). The findings of these ANOVAs (provided in Appendix N) supported those of the regressions here described.

M.1 Sample groupings

The strategic behaviour variable, which comprised the mean scores in response to the twenty-eight vocabulary-related strategic behaviour items in part 2 of the questionnaire, was taken as the independent variable of primary interest. In terms of the other two independent variables, the sample was divided into a three-level ordinal variable based on length of stay (group 1, 6-10 weeks; group 2, 11-15 weeks; group 3, 16-20 weeks) and into a binary categorical variable based on the location of the SAE (Oxford or London).

M.2 Parametric assumptions

Here follows a description of how the variables were investigated and treated to ensure they met the parametric assumptions required for regression analysis.

M.2a *Ratio of cases to independent variables*

As described in 4.3.2.2a, all three dependent variables (receptive, productive and overall vocabulary gain) exceeded the ratio of cases to independent variables advised by Tabachnick and Fidell (2001).

M.2b *Treatment of outliers*

A discussion of the treatment of outliers detected in the overall and receptive vocabulary gain variables is given in 4.1.1.4a and 4.1.1.4b respectively. An account of the logarithmic transformation of the productive vocabulary gain variable in response to outliers is provided in 4.1.1.4c. No outlaying cases existed in any of the independent variables.

M.2c *Multicollinearity and singularity*

Correlation analyses were conducted on the three independent variables to be used in this regression model to find one significant correlation: there was a positive relationship between location and identification with vocabulary-related strategic behaviour [r=.130,

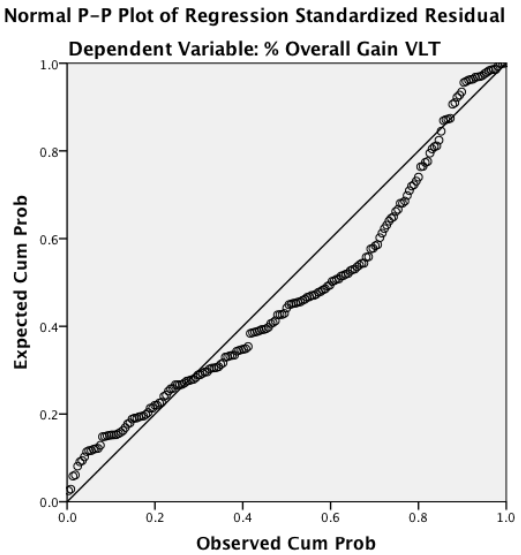
N=241, *p*=.044]. Berry (1993) states that r values above .9 indicate that independent variables are too highly correlated for use in a regression model, and Tabachnick and Fidell (2001) suggest that r values as low as .5 may be problematic. Therefore, the r value found in this correlation was deemed unproblematic. Moreover, the tolerance values for the three independent variables as shown in the collinearity statistics output of the initial regression model (described in K.2d) were all high at .970 and above. These explorations provided strong evidence to suggest that the assumption of the absence of multicollinearity had not been violated.

M.2d *Normality, linearity and homoscedasticity of residuals*

An initial regression was run to check the assumptions of normality, linearity and homoscedasticity. The initial regression model took overall vocabulary gain as the dependent variable, the continuous strategic behaviour variable as the independent variable of primary interest (entered as block 1), and location (as a binary categorical variable) and length of stay (as the two dummy variables described in 4.3.2.2d) as the secondary independent variables, entered as block 2.

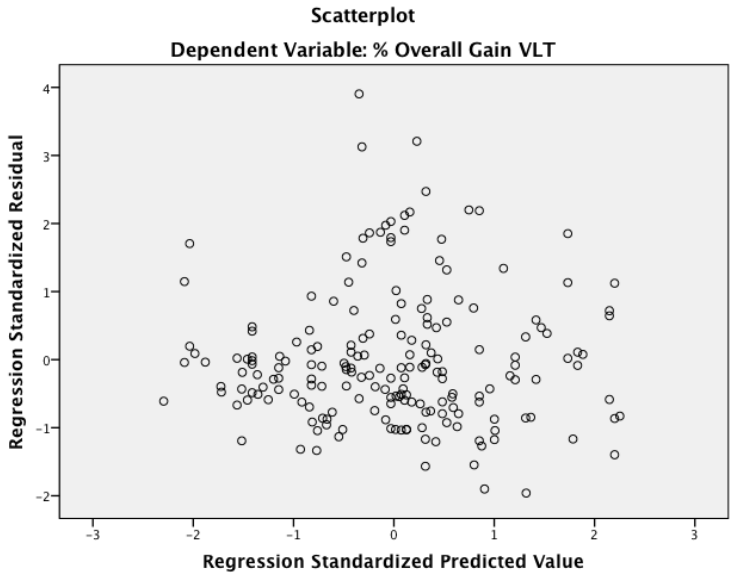
The normal probability plot (figure M.1) exhibited a reasonably straight, diagonal linear manifestation of the points, which suggested that there was no major deviation from the assumption of normality.

Figure M.1: Initial regression normal probability plot



Examination of the residuals scatterplot (figure M.2) for the model indicated that there was very little deviation from normality and that heteroscedasticity and non-linearity of residuals were not an issue.

Figure M.2: Initial regression residuals scatterplot



M.2e *Independence of errors*

The Durbin-Watson statistic of the final model of the initial regression was 2.18, suggesting that the residuals were uncorrelated and therefore that the assumption of independence of errors had not been violated.

M.2f *Solution outliers*

In order to inspect the model for solution outliers, the Mahalanobis distances produced by the initial regression were investigated. Tabachnick and Fidell (2001) state that the critical chi-square value for a model with four independent variables is 18.47 and that any cases that exceed this value can be deemed as solution outliers. No cases in the model had a Mahalanobis distance value exceeding 18.47: the minimum distance was 1.56 and the maximum 15.67 (M=3.98, SD=2.07).

M.3 Receptive vocabulary gain

A hierarchical multiple regression analysis was conducted to determine the extent to which the independent variables (block 1=identification with vocabulary-related strategic behaviour; block 2=length of stay and location of SAE) were able to predict the receptive vocabulary gain scores. After the entry of block 1, the proportion of variance explained was 5.9% and identification with vocabulary-related strategic behaviour was shown to have a significant effect on receptive vocabulary gain [F(1, 194)=12.13, p=.001]. After block 2 had been entered into the model, the model as a whole explained 7.9% of the variance and this model was also significant [F(4, 191)=4.08, p=.003]. However, the R Square change value was non-significant (.02, p=.25), which suggested that when the effect of strategic behaviour was controlled for, the variables in block 2 had no significant explanatory power.

Unsurprisingly, strategic behaviour was the only predictor variable to make a statistically significant contribution to the regression model (β=.226, p=.002): this finding mirrors that of the ANOVA of these variables (reported in Appendix N), in which strategic behaviour was found to have a significant main effect [F(1, 184)=9.849, p=.002] with a small effect size (Eta squared=.051). The positive beta value of the regression analysis finding indicates that, when the other independent variables were controlled for, a decrease in identification with the vocabulary-related strategies (which was manifested as a higher score on strategic behaviour because 5=This is not at all true of me) predicted an increase in receptive vocabulary gain.

Table M.1: Predictor variable contribution to receptive vocabulary gain regression equation

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-27.574	12.604		-2.188	.030
Questionnaire part 2 mean scores	.576	.165	.243	3.483	.001
2 (Constant)	-20.468	13.096		-1.563	.120
Questionnaire part 2 mean scores	.535	.170	.226	3.156	.002
Location of SAE	-2.732	4.913	-.039	-.556	.579
Length of stay dummy variable 1: 6-10 wks?	-1.061	6.014	-.013	-.176	.860
Length of stay dummy variable 2: 11-15 wks?	-11.053	6.098	-.135	-1.812	.071

M.4 Productive vocabulary gain

A hierarchical regression analysis was conducted to determine the extent to which the primary and secondary independent variables were able to predict the logarithmically transformed productive vocabulary gain variable. After the identification with strategic behaviour variable was entered, the overall model explained 0% of the variance in productive vocabulary gain and this finding was not statistically significant [F(1, 127)=.022, p=.883]. After block 2 (location and length of stay) had been entered, the model as a whole explained 1.9% of the variance but again, this model did not reach statistical significance [F(4, 124)=.590, p=.671]. Neither was the R square change statistic significant (.18, p=508). None of the predictor variables made a statistically significant contribution to the model. These findings were replicated in the corresponding ANOVA (Appendix N).

M.5 Overall vocabulary gain

A regression was also conducted to determine the extent to which the independent variables were able to predict overall vocabulary gain. After block 1 had been entered, the

model explained 5.3% of the variance. After block 2 had been entered, the model as a whole explained 8.3%. Both model 1 [F(1,191)=10.729, *p*=.001] and model 2 [F(4, 188)=4.249, *p*=.003] were statistically significant. However, similarly to the model that took receptive vocabulary gain as the dependent variable, the R square change value in this regression analysis was not significant (.030, *p*=.111), suggesting that the secondary independent variables failed to add any explanatory power to the model.

Two predictor variables made a significant contribution to the overall model: identification with vocabulary related strategic behaviour (*β*=.218, *p*=.003) and length of stay (*β*=-.161, *p*=.032). The corresponding ANOVA of these variables (Appendix N) found a main effect for strategic behaviour [F(1, 181)=15.45, *p*=.000], with a moderate effect size (Eta squared=.08), yet no significant main effect for length of stay [F(2, 181)=2.69, *p*=.07].

The positive beta weight of the strategic behaviour variable in this regression model suggested that a decrease in identification with the vocabulary-related strategies (which was manifested as a higher score on strategic behaviour because 5=This is not at all true of me) predicted an increase in overall vocabulary gain. The negative beta weight reported with reference to length of stay indicates that membership of the medium stay group (11-15 weeks) predicted lower overall vocabulary gain.

Table M.2: Predictor variable contribution to overall vocabulary gain regression equation

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-18.048	9.901		-1.823	.070
Questionnaire part 2 mean scores	.425	.130	.231	3.276	.001
2 (Constant)	-11.398	10.234		-1.114	.267
Questionnaire part 2 mean scores	.402	.133	.218	3.036	.003
Location of SAE	-2.858	3.840	-.053	-.744	.458
Length of stay dummy variable 1: 6-10 wks?	-5.335	4.700	-.083	-1.135	.258
Length of stay dummy variable 2: 11-15 wks?	-10.280	4.766	-.161	-2.157	.032

M.6 Summary of multiple regression analyses

The findings of the regression analyses here discussed indicate that the model was able to predict a small percentage of the variance in receptive and overall vocabulary gain. For receptive vocabulary gain, only strategic behaviour made a statistically significant contribution to the model, with a decrease in identification with vocabulary-related strategies predicting an increase in receptive vocabulary gain. For productive vocabulary gain, the model was unable to reveal any explanatory power. In terms of overall vocabulary gain, both strategic behaviour and length of stay made a significant impact.

Once more, stronger identification with vocabulary-related strategies predicted a decrease in gain. Additionally, membership of the medium stay group predicted lower overall gain.

Appendix N: ANOVA analyses of strategic behaviour, length of stay and location on vocabulary gain

A trio of three-way ANOVAs was conducted to explore the impact of each of the independent variables on receptive, productive and overall vocabulary gain respectively. The primary independent variable, identification with vocabulary-related strategic behaviour, was represented as binary ordinal variable dichotomised at the median split. Length of stay was a 3-level ordinal variable (short, medium or long), and location of SAE a dichotomous categorical variable (Oxford or London).

N.1 Receptive vocabulary gain

The first ANOVA of the series took receptive vocabulary gain as the dependent variable. The value of the Levene's test of equality of error variances was significant at p=.009, which highlighted that the variance in receptive vocabulary gain across the groups was unequal: in light of this finding, a more stringent significance level (.01) was adopted for evaluation of the findings of the ANOVA. The ANOVA found a statistically significant main effect for identification with vocabulary-related strategic behaviour [$F(1, 184)=9.849$, p=.002] with a small effect size (Eta squared=.051). No other significant main effects were detected at either $p<.01$ or $p<.05$.

N.2 Productive vocabulary gain

The ANOVA conducted with the logarithmically transformed productive vocabulary gain variable as the dependent variable found no statistically significant effect for any of the independent variables.

N.3 Overall vocabulary gain

The final ANOVA of the series took overall vocabulary gain as the dependent variable. The value of the Levene's test of equality of error variances was significant at p=.000, suggesting that the variance in overall vocabulary gain across the groups was unequal: once more, in light of this finding, a more stringent significance level (.01) was adopted. There was statistically significant main effect for strategic behaviour [$F(1, 181)=15.447$, p=.000], with a moderate effect size (Eta squared=.08). No other significant main effects were detected at either $p<.01$ or $p<.05$.

Appendix O: Regression analyses with mean informal contact and factored strategic behaviour

This variation of the regression model took the singular informal contact variable and the factored strategic behaviour variables as block 1, and location and the length of stay dummy variables as block 2.

O.1 Sample groupings

Informal L2 contact was represented in the present model by the mean scores of the items in part 1 of the questionnaire. Strategic behaviour was represented by the five factor variables. In terms of the secondary independent variables, the sample was divided into a three-level ordinal variable based on length of stay and into a binary categorical variable based on location.

O.2 Parametric assumptions

Here follows a description of how the variables were investigated and treated to ensure they met the parametric assumptions required for regression analysis.

O.2a *Ratio of cases to independent variables*

As described in 4.3.2.2a, all three dependent variables (receptive, productive and overall vocabulary gain) exceeded the ratio of cases to independent variables advised by Tabachnick and Fidell (2001).

O.2b *Treatment of outliers*

A discussion of the treatment of outliers detected in the overall and receptive vocabulary gain variables is given in 4.1.1.4a and 4.1.1.4b respectively. An account of the logarithmic transformation of the productive vocabulary gain variable in response to outliers is provided in 4.1.1.4c. No outlaying cases existed in any of the independent variables.

O.2c *Multicollinearity and singularity*

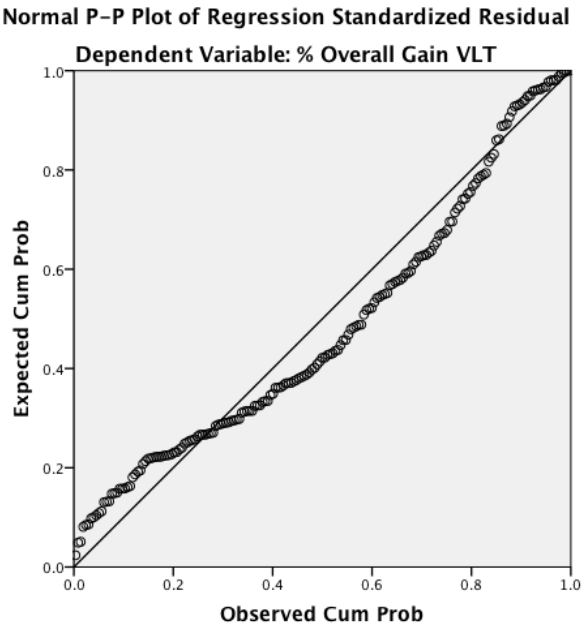
An initial regression was run to check the assumptions of multicollinearity, normality, linearity and homoscedasticity. The initial regression model took overall vocabulary gain as the dependent variable, the singular informal contact and factored strategic behaviour variables as block 1, and location and length of stay as block 2.

The tolerance values for the independent variables as shown in the collinearity statistics output of the initial regression were all high (.529 to .881). Moreover, inspection of the correlations output revealed that none of the variables had a bivariate correlation higher than .56: Tabachnick & Fidell (1996) suggest that correlations above .7 are too high for inclusion of the variables in a regression model. Therefore, these figures provided strong evidence to suggest that the assumption of the absence of multicollinearity had not been violated.

O.2d *Normality, linearity and homoscedasticity of residuals*

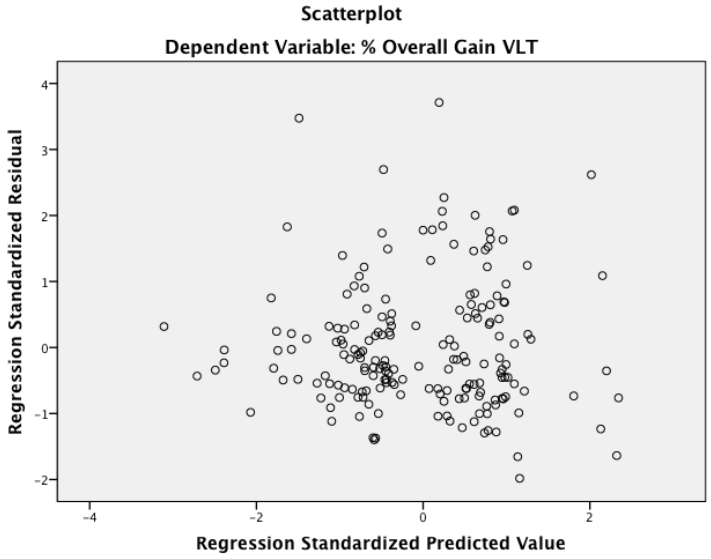
The normal probability plot (figure O.1) exhibited a largely straight, diagonal linear manifestation of the residuals with a slight curve to the right, suggesting that the assumption of normality was largely intact.

Figure O.1: Initial regression normal probability plot



Examination of the residuals scatterplot (figure O.2) indicated that there was a slight deviation from normality (in that slightly more residuals could be observed above the centre of the scatterplot than below) but that heteroscedasticity and non-linearity of residuals were not an issue.

Figure O.2: Initial regression residuals scatterplot



O.2e *Independence of errors*

The Durbin-Watson statistic of the final model of the initial regression was 2.02, suggesting that the residuals were uncorrelated and therefore that the assumption of independence of errors had not been violated.

O.2f *Solution outliers*

In order to inspect the model for solution outliers, the Mahalanobis distances produced by the initial regression were investigated. Tabachnick and Fidell (2001) state that the critical chi-square value for a model with nine independent variables is 27.88 and that any cases that exceed this value can be deemed as solution outliers. There were nine independent variables in the model because strategic behaviour comprised five factor variables and length of stay comprised two dummy variables. No cases in the model had a Mahalanobis distance value exceeding 27.88: the minimum distance was 3.06 and the maximum 20.74 (M=8.95, SD=3.09).

O.3 Receptive vocabulary gain

A hierarchical multiple regression analysis was conducted to determine the extent to which the independent variables (block 1=identification with informal L2 contact scenarios and identification with factored strategic behaviour; block 2=length of stay and location of SAE) were able to predict the receptive vocabulary gain scores. After the entry of block 1, the proportion of variance explained was 6.5% and identification with informal contact and factored strategic behaviour was shown to have a significant effect on receptive vocabulary gain [F(6, 188)=2.18, p=.047]. After block 2 had been entered into the model, the model as a whole explained 10.2% of the variance and this model was also significant [F(9, 185)=2.35, p=.016]. The R Square change value was non-significant (.037, p=.056), which suggested that when the effect of identification with informal contact and factored strategic behaviour was controlled for, the variables in block 2 had no significant explanatory power.

Three predictor variables made a statistically significant contribution to the regression model: identification with Appraisal strategies (β=.237, p=.011), identification with Establishment strategies (β=-.156, p=.044) and length of stay (β=-.242, p=.008). The corresponding ANOVA (reported in Appendix P) found a statistically significant main effect for identification with vocabulary-related strategic behaviour [F(1, 173)=7.53, p=.007] with a small to medium effect size (Eta squared=.042), but failed to find a significant effect for length of stay.

The positive beta value of the finding pertaining to Appraisal strategies indicates that, when the other independent variables were controlled for, an increase in identification with the Appraisal strategies predicted an increase in receptive vocabulary gain. The negative beta values of the findings pertaining to Establishment strategies and length of stay revealed that when the other variables were controlled for, stronger identification with

Establishment strategies and belonging to the medium stay group were both predictors of lower receptive gain.

Table O.1: Predictor variable contribution to receptive vocabulary gain regression equation

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	17.297	18.313		.945	.346
Cognitive Action	1.141.	.444	.210	2.568	.011
Appraisal	.776	.548	.114	1.415	.159
Visual Reference	1.277	.976	-.100	-1.308	.192
Establishment	-1.143	.573	-.152	-1.993	.048
Context Provision	.449	.862	.041	.516	.607
Questionnaire part 1 mean scores	-.170	.203	-.066	-.816	.404
2 (Constant)	27.334	18.643		1.466	.144
Cognitive Action	.425	.519	.078	.819	.414
Appraisal	1.618	.630	.237	2.567	.011
Visual Reference	-1.888	.992	-.148	-1.903	.059
Establishment	-1.173	.579	-.156	-2.026	.044
Context Provision	.677	.881	.062	.768	.443
Questionnaire part 1 mean scores	-.182	.207	-.071	-.881	.379
Location of SAE	-2.191	5.147	-.032	-.426	.671
Length of stay dummy variable 1: 6-10 wks?	.780	6.597	.009	.118	.906
Length of stay dummy variable 2: 11-15 wks?	-19.783	7.331	-.242	-2.698	.008

O.4 Productive vocabulary gain

A hierarchical regression was also conducted to determine the extent to which the primary and secondary independent variables were able to predict the logarithmically transformed productive vocabulary gain variable. After the identification with informal contact and factored strategic behaviour variables were entered, the overall model explained 7.5% of the variance in the productive gain scores. After location and length of stay had been entered, the overall model explained 9.9%. Neither model 1 [F(6, 122)=1.64, p=.142] nor model 2 [F(9, 119)=1.45, p=.173] was statistically significant. This finding was supported by the corresponding ANOVA (reported in Appendix N).

O.5 Overall vocabulary gain

In the final regression model the dependent variable was overall vocabulary gain. After block 1 had been entered, the model explained 5.9% of the variance in overall gain, and after block 2 had been entered the model as a whole explained 9.3%.

Model 1 was not statistically significant [F(6, 185)=1.95, p=.075], but model 2 was [F(9, 182)=2.09, p=.033]. The R square change statistic was not significant (.034, p=.081), which indicated that location and length of stay as a block added no explanatory power to the model when identification with informal contact and strategic behaviour were controlled for.

Only length of stay made a significant contribution to the overall model (β=-.220, p=.016): belonging to the medium stay group predicted lower overall vocabulary gain. This finding was supported by the corresponding ANOVA reported in Appendix P, although the ANOVA also revealed a main effect of strategic behaviour.

Table O.2: Predictor variable contribution to overall vocabulary gain regression equation

Coefficients					
Model	Unstandardised coefficients		Standardised coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	16.406	14.389		1.140	.256
Cognitive Action	1.039	.349	.246	2.976	.003
Appraisal	.327	.431	.062	.759	.449
Visual Reference	-1.015	.767	-.102	-1.323	.187
Establishment	-.685	.450	-.118	-1.521	.130
Context Provision	-.116	.685	-.014	-.170	.865
Questionnaire part 1 mean scores	-.088	.159	-.044	-.551	.582
2 (Constant)	25.554	14.679		1.741	.083
Cognitive Action	.665	.409	.158	1.627	.105
Appraisal	.797	.496	.150	1.606	.110
Visual Reference	-1.472	.781	-.149	-1.885	.061
Establishment	-.601	.456	-.103	-1.317	.189
Context Provision	.059	.694	.007	.085	.933
Questionnaire part 1 mean scores	-.142	.163	-.071	-.870	.385
Location of SAE	-1.485	4.053	-.028	-.366	.715
Length of stay dummy variable 1: 6-10 wks?	-5.832	5.194	-.091	-1.123	.263
Length of stay dummy variable 2: 11-15 wks?	-14.035	5.772	-.220	-2.431	.016

O.6 Summary of multiple regression analyses

The findings of these regression analyses indicate that the final model was able to predict a small percentage of receptive and overall vocabulary gain. For receptive vocabulary gain, strategic behaviour and length of stay made a statistically significant contribution to the model, with a increase in identification wit Appraisal strategies predicting an increase in overall gain, and a decrease in identification with Establishment strategies and belonging to the medium stay group predicting an decrease in receptive vocabulary gain. For overall gain, length of stay was the sole significant contributor, with belonging to the medium stay group predicting lower overall vocabulary gain. The model failed to predict gains in productive vocabulary knowledge.

Appendix P: ANOVA analyses of informal contact, strategic behaviour, length of stay and location on vocabulary gain

A triumvirate of four-way ANOVAs was conducted to explore the impact of the four independent variables in the study on receptive, productive and overall vocabulary gain respectively. The primary independent variables, identification with informal L2 contact scenarios and identification with vocabulary-related strategic behaviour, were represented as binary ordinal variables dichotomised at the median split. Length of stay was a 3-level ordinal variable (short, medium or long), and location of SAE a dichotomous categorical variable (Oxford or London).

P.1 Receptive vocabulary gain

The first ANOVA of the series took receptive vocabulary gain as the dependent variable. The value of the Levene's test of equality of error variances was significant at p=.020, which highlighted that the variance in receptive vocabulary gain across the groups was unequal: in light of this finding, a more stringent significance level (.01) was adopted for evaluation of the findings of the ANOVA. The ANOVA found a statistically significant main effect for identification with vocabulary-related strategic behaviour [F(1, 173)=7.53, p=.007] with a small effect size (Eta squared=.042). No other significant main effects were detected at either p<.01 or p<.05.

P.2 Productive vocabulary gain

The second ANOVA took the logarithmically transformed productive gain variable as the dependent variable. The value of the Levene's test of equality of error variances was significant at p=.046 and therefore a more significance level of p=.01 was adopted for evaluation of the findings. There was no statistically significant effect for any of the independent variables at the p=.01 level. Identification with informal contact had an effect that would have reached significance had the variance in productive gain across the groups been equal [F(1, 108)=3.93, p=.05].

P.3 Overall vocabulary gain

The final ANOVA of the series took overall gain as the dependent variable. Levene's test of equality of error variances was significant (p=.000) and as such p<.01 was adopted as the significance level. However, at this level only identification with strategic behaviour had a significant main effect [F(1, 170)=9.94, p=.002], with a medium effect size (Eta squared=.06). The effect of length of stay was significant at p<.03 [F(2, 170)=3.67, p=.027] with a small effect size (Eta squared=.041). Post-hoc comparisons using the Scheffe test indicated that the mean vocabulary gain of the medium stay group (11-15 weeks; N=47, M=5.22, SD=21.85) was significantly different at the p<.03 level from the mean gain of the long stay group (16-20 weeks; N=101, M=18.04, SD=27.73; p=.022). The short stay group did not differ significantly from either other group.