

**“WORKING OUT” THE RELATIONSHIP
BETWEEN INFORMAL LANGUAGE CONTACT AND
PHRASAL VERB DEVELOPMENT OF
INTERNATIONAL STUDENTS IN THE UK**



Siyang Zhou

KELLOGG COLLEGE

THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

2022

Table of Contents

Abstract.....	vi
Acknowledgments.....	vii
List of Tables.....	ix
List of Figures.....	xi
List of Abbreviations.....	xiii
CHAPTER I INTRODUCTION.....	1
1.1 Background to the study	1
1.2 The study.....	2
1.2.1 Research questions.....	3
1.3 Rationale and significance of the study	4
1.4 Defining key concepts.....	6
1.4.1 Formulaic language.....	6
1.4.2 Phrasal verbs	7
1.4.3 Study-abroad experience.....	7
1.4.4 Language contact	8
1.4.5 Social networks.....	8
1.5 Theoretical frameworks	8
1.6 Thesis outline.....	10
CHAPTER II LITERATURE REVIEW	12
2.1 Study abroad research.....	12
2.1.1 Linguistic and non-linguistic outcomes.....	13
2.1.2 Vocabulary development and language contact	15
2.1.3 Variability in the impact of study-abroad on SLA.....	18
2.2 Researching phrasal verbs.....	21
2.2.1 Vocabulary learning in SLA.....	21
2.2.2 Acquisition of formulaic language.....	22
2.2.3 Research on phrasal verbs.....	24
2.2.4 The assessment of phrasal verb knowledge	28
2.2.5 Factors influencing phrasal verb knowledge	32
2.3 Informal language contact in SA	35
2.3.1 Language contact	35
2.3.1.1 Language contact research in SA.....	35

2.3.1.2 Measuring language contact during SA.....	37
2.3.2 Social networks	43
2.3.2.1 Social network research in SA.....	43
2.3.2.2 Measuring social networks during SA.....	46
2.3.2.3 Social network models for international students.....	52
2.4 Investigating phrasal verbs and informal language contact.....	56
2.4.1 Formulaic language development and language contact	56
2.4.2 Rationale for the current study and original contributions	60
CHAPTER III METHODOLOGY	63
3.1 Research questions.....	63
3.2 Study design.....	64
3.3 Sample.....	65
3.3.1 Sampling strategy.....	65
3.3.2 The foundation programs	67
3.3.3 Sample characteristics.....	68
3.3.4 Participant attrition.....	70
3.4 Research instruments	72
3.4.1 Standardized PV tests	72
3.4.1.1 Productive test.....	76
3.4.1.2 Receptive test	76
3.4.1.3 Validity and reliability	77
3.4.2 the Vocabulary Levels Test	79
3.4.3 Questionnaires.....	80
3.4.3.1 Language Contact Questionnaire	81
3.4.3.2 Demographic questionnaire	83
3.4.3.3 The Study Abroad Social Network Survey.....	84
3.4.4 IELTS speaking test	91
3.4.5 Semi-structured interview	92
3.5 Procedure	93
3.5.1 Study piloting.....	94
3.5.2 Time 1 data collection.....	96
3.5.3 Time 2 data collection.....	98
3.5.4 Time 3 data collection.....	98
3.6 Data analysis	99
3.6.1 Research Question 1	99
3.6.1.1 PV tests	99
3.6.1.2 PV learner corpus.....	101
3.6.2 Research Question 2	102

3.6.2.1 Pre-COVID data.....	103
3.6.2.2 Post-COVID data	106
3.6.3 Research Question 3	107
3.7 Methodological limitations	108
3.8 Ethical considerations	110
CHAPTER IV FINDINGS PART I.....	113
4.1 TEST DATA.....	113
4.1.1 Descriptives and normality	113
4.1.1.1 Time 1 performance	114
4.1.1.2 Time 2 performance	118
4.1.1.3 Time 3 performance	122
4.1.1.4 Phrasal verb gains	124
4.1.1.5 Summary of descriptives and normality	128
4.1.2 Comparison of means	129
4.1.2.1 Performance between types of PV knowledge	129
4.1.2.2 Vocabulary knowledge scores at Time 1 and Time 2.....	130
4.1.2.3 Phrasal verb knowledge scores at Time 1, 2, and 3.....	131
4.1.2.4 Summary of comparison of means	132
4.1.3 Non-exposure factors that influence the phrasal verb knowledge	133
4.1.4 Further analysis of phrasal verb tests.....	141
4.2 LEARNER CORPUS	145
4.2.1 Quantitative analysis of learner corpus.....	146
4.2.1.1 Number of phrasal verbs	146
4.2.1.2 Sophistication of phrasal verbs	149
4.2.1.3 Accuracy of phrasal verbs.....	151
4.2.1.4 Summary of quantitative analysis of phrasal verbs	153
4.2.2 Qualitative analysis of inaccurate PV use.....	153
4.2.3 Non-exposure factors that influence phrasal verb production	154
4.3 SUMMARY OF PHRASAL VERB DEVELOPMENT	157
CHAPTER V FINDINGS PART II.....	159
5.1 PRE-COVID LANGUAGE CONTACT	160
5.1.1 The Language Contact Questionnaire.....	160
5.1.1.1 Descriptives and normality	160
5.1.1.2 Comparison of means	164
5.1.1.3 Exploratory factor analysis	165
5.1.1.4 Factors influencing language contact.....	168
5.1.1.5 Summary of the LCQ data	172
5.1.2 The Study Abroad Social Network Questionnaire.....	172

5.1.2.1 Descriptives and normality	172
5.1.2.2 Comparison of means	176
5.1.2.3 Social network characteristics.....	176
5.1.2.4 Factors influencing social networks.....	178
5.1.2.5 Summary of the SASNS data.....	181
5.1.3 Interview data.....	182
5.1.3.1 Setting the scene	183
5.1.3.2 Language contact	184
5.1.3.3 Social networks.....	190
5.1.4 Summary of pre-COVID informal language contact.....	199
5.2 POST-COVID LANGUAGE CONTACT PATTERNS	199
5.2.1 The Language Contact Questionnaire.....	200
5.2.1.1 Time 3 data	200
5.2.1.2 Comparison of means	201
5.2.2 The Study Abroad Social Network Questionnaire.....	202
5.2.2.1 Time 3 data	202
5.2.2.2 Comparison of means	204
5.2.3 Interview data.....	205
5.2.3.1 Informal language contact.....	205
5.2.3.2 Social network	208
5.2.3.3 SA experience	210
5.2.4 Summary of the post-COVID informal language contact	212
5.3 SUMMARY OF INFORMAL LANGUAGE CONTACT.....	213
CHAPTER VI FINDINGS PART THREE	215
6.1 QUANTITATIVE ANALYSIS	215
6.1.1 Correlation	216
6.1.2 Multiple regression	219
6.1.3 Logistic mixed-effects modeling	224
6.1.4 Summary of quantitative findings.....	227
6.2 QUALITATIVE ANALYSIS	227
6.2.1 Attitudes toward phrasal verbs.....	228
6.2.2 Phrasal verb learning in study-abroad.....	232
6.2.3 Summary of qualitative findings.....	238
CHAPTER VII DISCUSSION	239
7.1 Phrasal verb development.....	239
7.1.1 Developmental profile	239
7.1.2 Influential factors	245
7.2 Informal language contact.....	247

7.2.1 Language contact patterns and developmental profile.....	247
7.2.1.1 Language contact	247
7.2.1.2 Social networks	250
7.2.2 Influential factors	255
7.2.3 The impact of COVID-19	256
7.2.3.1 Language contact	256
7.2.3.2 Social networks	260
7.2.3.3 SA experience	261
7.3 The relationship between phrasal verb and language contact.....	263
7.3.1 Usage-based approach to phrasal verb acquisition in SA.....	264
7.3.2 Language contact and PV development.....	266
7.3.3 Social networks and PV development	270
7.4 Summary of Discussion	272
CHAPTER VIII CONCLUSIONS	274
8.1 Key findings and contributions.....	274
8.2 Limitations	277
8.3 Implications.....	279
8.4 Recommendations for future research	281
8.5 Concluding remarks	282
References.....	284
Appendix A Ethics approval	309
Appendix B Research information sheet and consent form.....	310
Appendix C Participant Recruitment Presentation	312
Appendix D Test papers	314
Appendix E the Language Contact Questionnaire.....	318
Appendix F IELTS speaking test questions	319
Appendix G Interview schedules	322
Appendix H Quantitative analysis for Finding Chapter IV	324
Appendix I Quantitative analysis for Finding Chapter V	332
Appendix J Sample coding scheme of thematic analysis	342
Appendix K Quantitative analysis for Finding Chapter VI	346

Abstract

Study abroad (SA) has long been regarded as a learning environment that is beneficial to second language acquisition (SLA), especially in competencies related to social interactions (Kinging, 2011). However, little is known about whether SA students make actual gains in domains of English they are typically exposed to outside of classroom settings such as phrasal verbs (PVs), and whether their informal language contact contributes to such development. To explore this relationship, this study investigates students' PV development and informal language contact, to understand how they make use of the L2 exposure during SA.

This longitudinal mixed-methods study collected data from over 250 international foundation year students studying in the UK three times in one academic year. PV data were collected through standardized tests and speech samples, while L2 contact was measured via language contact questionnaires, social network surveys, and semi-structured interviews. Paired-samples t-test revealed that there were significant gains in receptive and productive PV scores between Time 1 and Time 2, while the overall informal language contact and social networks of the participants did not see significant changes from Time 1 to Time 2. Correlation and multiple regressions analyses revealed a stronger relationship between informal language contact and PV knowledge at Time 2 than at Time 1. Mixed-effects modeling revealed that the overall English proficiency of the participants, corpus frequency of the PVs in the British National Corpus, and the amount of language contact were significant predictors of PV scores, while semantic opacity of PVs, the Chinese L1 of the participants, and L2 social networks were not.

This study underscored that the degree-oriented SA participants may have different prioritization and leisure time routines compared with summer school participants or exchange students abroad. With academic prioritization, the participants made limited gains in colloquial L2 vocabulary competency and their highly academic L2 contact did not greatly contribute to PV development.

Acknowledgments

When I was 22, I buried a humble dream of doing a PhD at the best university in the world. Now, when I could not be closer to the end of this unforgettable intellectual journey, I want to express my gratitude to many people. Firstly, I would like to express my sincere gratitude to Prof. Heath Rose, Prof. Paul Meara, and Dr. Jessica Briggs Baffoe-Djan, who have supervised me at different stages. I am eternally grateful for all their comments, suggestions, and feedback. They are some of the brightest minds in the field and they are always my role models to look up to.

I also feel fortunate in meeting many great colleagues and academics in the department. Special thanks go to Prof. Victoria Murphy and Dr. Robert Woore for their insightful comments to my work and Dr. Elizabeth Wonnacott and Dr. Sophie Booton for their help on mixed-effects modeling. And I will not forget various knowledge exchanges with other external scholars who have inspired me in different ways.

I gratefully acknowledge the help from all the participants and teachers that were involved in my study. Especially, I am deeply indebted to Mr. James Martin, Mr. Amarpal Kahlwan, and Ms. Maria Calonico, for facilitating the longitudinal project. I am truly thankful to all the teachers and hundreds of students in Oxford and China that participated in the piloting of this study. Further, I am beyond grateful for the technical help from my Oxford friend, John Fraser, for developing an online instrument for this study. My gratitude also goes to the generous support from the China Scholarship Council to fund my study and to Kellogg College and the Education Department for the financial stipends for my academic trips.

I will not forget my dear friends in different corners of the world: Xiaoxin, István, Leila, Fish, Weichen, Jamie, and Zhanxin for always being by my side. I would like to express my deepest appreciation to my parents, who have always given me unconditional love, support, and freedom to pursue my dreams. Additionally, I'd like to thank Yaohua for his unconditional love, sweet company, and immense support during my remote PhD

days. Lastly, some ups and downs happened in my life during the PhD: going through an unprecedented pandemic, embracing the supervisor change, living with and conquering some health issues... I believe these experiences have helped me become a stronger person and I am ready to take up any other challenges in the future with resilience and persistence.

List of Tables

Table 2.1 Comparison of existing social network questionnaires	51
Table 3.1 Simultaneous mixed-methods study design	64
Table 3.2 Research questions and corresponding data	65
Table 3.3: Participant number across sites	67
Table 3.4: Nationalities of participants	68
Table 3.5: Interviewees profile	69
Table 3.6: Sample size for each quantitative instrument	71
Table 3.7: PV target items in the tests, ranked by BNC frequency	74
Table 3.8 Sample item in the productive PV test	76
Table 3.9 Sample item in the receptive PV test	77
Table 3.10 Internal consistency of receptive and productive PV tests	78
Table 3.11 Example VLT item	80
Table 3.12 IELTS speaking test topics	92
Table 3.13 Coding protocol of PVs in the speech samples	101
Table 3.14 Coding protocol of SASNS network characteristics	103
Table 3.15 Analysis technique for RQ2a	105
Table 4.1a Summary of descriptives of each test from all participants at each timepoint	114
Table 4.1b Summary of normality check of each test	114
Table 4.2: Descriptives of PV item scores at T1	117
Table 4.3: Descriptives of PV item scores at T2	120
Table 4.4: Descriptives of PV item scores at T3	124
Table 4.5 Summary of PV gains between two time points	125
Table 4.6 Comparison of means between productive and receptive PV tests	130
Table 4.7 Tests with significant gains over time points after Bonferroni correction	132
Table 4.8 Summary of non-language-contact factors analysis techniques	133
Table 4.9 Spearman correlation between corpus frequency and PV scores	135
Table 4.10 Pearson correlations between overall English proficiency and PV scores	137
Table 4.11: Independent-samples <i>t</i> -tests in vocabulary scores between learner groups	138
Table 4.12 Type of changes of PV knowledge	143
Table 4.13 Types of gains of PV knowledge	143
Table 4.14 Overview of PV type number in the learner corpus	146
Table 4.15 Most frequent PVs from the whole learner corpus	147
Table 4.16 Descriptive statistics of PV number	147
Table 4.17 Descriptive of sophistication of PVs in the learner corpus	150

Table 4.18a Descriptives of semantic accuracy of PVs in the learner corpus	151
Table 4.18b Descriptives of grammatical accuracy of PVs in the learner corpus	151
Table 4.19 Coding scheme for semantic errors	153
Table 4.20 Coding scheme for grammatical errors	154
Table 4.21 Summary of factors investigated and statistical techniques	155
Table 5.1 Descriptive statistics of T1 and T2 LCQ breakdown	162
Table 5.2 Descriptive statistics of the T1 – T2 LCQ item gains	163
Table 5.3 Comparison of LCQ items between T1 and T2	165
Table 5.4 Pattern Matrix of T1 LCQ data	166
Table 5.5 Pattern Matrix of T2 LCQ data	167
Table 5.6 Significant L1 difference in LCQ scores	169
Table 5.7 Correlation between overall English proficiency and LCQ scores	169
Table 5.8 Correlation between motivation and LCQ scores	170
Table 5.9 Descriptive statistics of network variables in the SASNS	173
Table 5.10 T1 to T2 gains of the SASNS	175
Table 5.11a Geographic distance of listed friends	177
Table 5.11b Cultural distance of listed friends	177
Table 5.12 The first encounter of the listed English-speaking friends in the SASNS	178
Table 5.13 Friendship intensity in the SASNS	178
Table 5.14 Friends' English proficiency in the SASNS	178
Table 5.15a L1 difference in the T1 SASNS variables	179
Table 5.15b L1 difference in the T2 SASNS variables	179
Table 5.16 Interviewee profile by LCP scores	182
Table 5.17 Descriptive statistics of the T3 LCQ breakdown	201
Table 5.18 One-way repeated-measures ANOVA of LCQ items across 3 time points	202
Table 5.19 Descriptive statistics of T3 SASNS	203
Table 5.20 Factors that influence contact frequency in the lockdown	208
Table 6.1 Correlation between T1 PV scores and LCQ factors	216
Table 6.2 Correlation between T1 PV scores and LCQ items	217
Table 6.3 Correlation between T2 PV scores and LCQ factors	218
Table 6.4 Correlation between T2 PV scores and LCQ items	218
Table 6.5 Hierarchical regression results of receptive PV scores at T2	222
Table 6.6 Hierarchical regression results of productive PV scores at T2	223
Table 6.7 Model summary of the mixed-effects model for receptive PV scores	226
Table 6.8 Model summary of the mixed-effects model for productive PV scores	226

List of Figures

Figure 2.1 Productive test and receptive test in Schmitt and Redwood (2011)	28
Figure 2.2 Receptive PV test in Kamarudin et al. (2019)	32
Figure 2.3 Part of the LCP	38
Figure 2.4 The language log (Martinsen et al., 2010)	40
Figure 2.5 Part of Block 1 of the DLQ	42
Figure 2.6 Example of ego-net	44
Figure 2.7 Screenshot of the SNQ	48
Figure 2.8 Part of the SNSS	49
Figure 2.9 Screenshot of the personal network questionnaire	50
Figure 2.10 The concentric framework of Coleman (2013)	53
Figure 2.11 Kachrus' Three Circles of English (1982)	55
Figure 2.12 SA social network model in the UK	56
Figure 3.1 Liaison hierarchy of the data collection	66
Figure 3.2. Reasons for participant attrition at T1 and T2	72
Figure 3.3 Sample item in the LCQ	81
Figure 3.4 Structure of the SASNS and reference instruments for each part	85
Figure 3.5a Part 1 of SASNS	86
Figure 3.5b Part 2 of SASNS	88
Figure 3.5c Part 3 of SASNS	89
Figure 3.5d Part 4 of SASNS	90
Figure 3.5e Part 5 of SASNS	90
Figure 3.6 Study timeline	94
Figure 3.7a Procedure of T1 data collection	96
Figure 3.7b Procedure of T1 data collection	98
Figure 3.7c Procedure of T3 data collection	98
Figure 3.8 Quantitative analysis to answer RQ 3	108
Figure 4.1: Histogram of T1 receptive PV scores	115
Figure 4.2: Histogram of T1 productive PV scores	116
Figure 4.3: Histogram of T1 VLT scores	118
Figure 4.4: Histogram of T2 receptive PV scores	119
Figure 4.5: Histogram of T2 productive PV scores	120
Figure 4.6: Histogram of T2 VLT scores	121
Figure 4.7: Histogram of T3 receptive PV scores	122
Figure 4.8a: Histogram of T3 productive PV scores	123
Figure 4.8b: Histogram of transformed T3 productive PV scores	123
Figure 4.9 Histogram of T1 to T2 receptive PV test gains	126

Figure 4.10 Histogram of T1 to T2 productive PV test gains	126
Figure 4.11 Spaghetti plot of T1, 2, 3 receptive test scores	127
Figure 4.12 Spaghetti plot of T1, 2, 3 productive test scores	128
Figure 4.13a Receptive PV mean scores trend by descending corpus frequency order	136
Figure 4.13b Productive PV mean scores trend by descending corpus frequency order	136
Figure 4.14 Coding of the PV knowledge level	141
Figure 4.15 Coding of the developmental profile of PV knowledge	141
Figure 4.16 Degree of knowledge of target PVs	142
Figure 4.17 Individual changing trends of PV knowledge	143
Figure 4.18a PV number change between T1 and T2 by participant (N=15)	149
Figure 4.18b PV number change over T1, T2, and T3 by participant (N=9)	149
Figure 4.19 Developmental profile of PV sophistication of each part	150
Figure 4.20a Developmental profile of PV semantic accuracy of each part	152
Figure 4.20b Developmental profile of PV grammatical accuracy of each part	152
Figure 5.1 Histogram of T1 LCQ scores	161
Figure 5.2 Histogram of T2 LCQ scores	161
Figure 5.3 Histogram of T1 to T2 LCQ gains	163
Figure 5.4 Scree plot of T1 LCQ EFA	165
Figure 5.5a SASNS activity frequencies by rank	174
Figure 5.5b SASNS topic frequencies by rank	174
Figure 5.6 Histogram of T3 LCQ scores	200
Figure 5.7a T2 SASNS activity frequencies by rank	204
Figure 5.7b T3 SASNS topic frequencies by rank	204
Figure 6.1 Two models for multiple regression	220
Figure 7.1 The social networks model for NNS students in the UK	253
Figure 7.2 Reasons for low L2 contact during the lockdown	258
Figure 7.3 A usage-based conceptual framework of PV acquisition during SA	264

List of Abbreviations

ANOVA	Analysis of variance
BNC	British National Corpus
COCA	Corpus of Contemporary American English
COVID-19	Corona Virus Disease 2019
CUREC	Central University Research Ethics Committee
DLQ	Daily Linguistic Questionnaire
EFL	English as a Foreign Language
EMAS Corpus	English of Malaysian Students Corpus
EMI	English-Medium-Instruction
ERASMUS	European Community Action Scheme for the Mobility of University Students
ESL	English as a Second Language
IELTS	International English Language Testing System
L1	First language
L2	Second language
L3	Third language
LANGSNAP	Social networks, target language interaction, and second language acquisition during the year abroad: A longitudinal study
LCP	Language Contact Profile
LCQ	Language Contact Questionnaire
LEQ	Language Engagement Questionnaire
LoS	Lengths of stay
M	Mean
MANCOVA	Multivariate analysis of covariance
MPT	Multi-word Phrases Test
OPI	Oral Proficiency Interview
PHaVe List	The phrasal verb Pedagogical List
PV	Phrasal verb
RQ	Research question
SA	Study abroad
SALA	Study Abroad and Language Acquisition
SASIQ	Study Abroad Social Interaction Questionnaire
SASNS	Study Abroad Social Network Survey
SD	Standard deviation
SLA	Second language acquisition

SNQ	Social Networks Questionnaire
SNSS	Social Network Strength Scale
T1	Time 1
T2	Time 2
T3	Time 3
VLT	Vocabulary Levels Test

CHAPTER I INTRODUCTION

1.1 Background to the study

The internationalization of higher education in recent decades has boosted the global phenomenon of study abroad (SA), as students have become more internationally mobile. Within the population of SA students, Chinese students in the UK represent a massive learner cohort. According to Study International (2018), the UK continues to be ranked as the second most popular SA destination for international students around the world, following the US. Among all countries of origin, China sends the largest number of students overseas every year (around 700,000), followed by India (189,500) and Korea (123,700) (WorldAtlas, 2017). Over 120,000 Chinese students were enrolled in British universities in the UK in 2018/2019, making up 35% of all non-EU students (Study in UK, 2020). In other words, one out of three non-European international students studying in the UK comes from China. Arguably, the strong presence of Chinese students in British universities is likely to continue, given fewer enrolments from EU students after Brexit (Morrison, 2021) and the geopolitical tension between China and the US (Lou, 2021).

SA requires a substantial investment of finances and time from the students and their parents—driven, in part, by an expectation of English language development while abroad. Thus, it is of both practical importance and theoretical significance to investigate the quality of the SA experience in terms of language use and social engagement and the extent to which the L2 proficiency of learners improves during SA. Related findings will not only help students and parents to weigh the realistic outcomes of SA in decision making but can also inform SA program coordinators about the extracurricular life of their students, which might contribute to better support provision for students.

This study chooses phrasal verbs (PVs), a niche area of L2 competence, for

investigation. This linguistic feature is useful for research in the SA context because formal classroom instruction often overlooks PV teaching, and idiomatic language expressions (including PVs) are usually acquired in daily life (Garnier & Schmitt, 2016). PVs, therefore, may provide a window into colloquial vocabulary development via informal language contact while undertaking SA.

Understanding local people's daily conversations is crucial for students' social adaptation and integration. Thus, the present study aims to investigate PV development among international students in the UK under the influence of informal language contact. The sample of the study was foundation year students in the UK, who planned to reside in the UK for their full degree (three years for bachelor's degrees; one or two years for masters' degrees). The findings from this study will not only shed light on the development of an underexplored L2 competence during SA but also reveal the extent to which foundation students are exposed to the English-speaking environment in their daily life, and whether L2 exposure in SA facilitates their development of English formulaic language.

1.2 The study

Given the mismatch between the genre of English that students are exposed to in class and their real-life colloquial communicative needs, this study aims to investigate whether the characteristics of their out-of-class L2 contact contributes to students' gain in PV knowledge. The first two aims of the study are to identify the PV developmental trajectory and informal language pattern changes of the participants in one academic year. The third aim of this study is to explore the relationship between these two constructs, to probe if this SA experience benefits such formulaic language improvement of the learners.

A longitudinal mixed-methods study collected data from over 250 international foundation year students studying in the UK for one academic year between September 2019 and June 2020. Data collection took place once per term and three times in total

(November 2019, February 2020, and May 2020). Two threads of data were gathered and analyzed: 1) PV data were collected through standardized tests and speech samples; 2) L2 contact was measured via language contact questionnaires, social network surveys, and semi-structured interviews. Other confounders, such as initial vocabulary knowledge and personal backgrounds (including motivation), were also assessed at the beginning of their stay to better investigate the relationship between PV command and language contact of L2 learners. Statistical comparisons of means were conducted to reveal whether there were significant gains in PV scores and significant changes in the language contact. Correlation analysis and between-groups comparison techniques were employed to identify non-exposure factors that influence the PV performance and language contact of the participants. Correlations, multiple regressions and mixed-effects modeling were performed to investigate the relationship between PV scores and language contact.

It should be noted that due to the COVID-19 outbreak, there were major alterations to the data collection arrangement at Time 3. The sudden halt of all social activities during the national lockdown deprived the participant of a normal SA experience. Although the original research purpose was partly unfulfilled, this study provided an opportunity to capture the impact of the external environment on L2 exposure of the international student body during the pandemic. This provides valuable first-hand insight into their SA experience during the COVID-19 lockdown.

1.2.1 Research questions

Within this context, the study answers the following research questions (RQs):

1. What is the developmental profile of PV knowledge among international foundation program students in the UK over one academic year?
 - 1a. What non-exposure factors influence the PV scores of the participants?
2. What are the patterns and developmental profile of informal L2 contact among international foundation program students in the UK over the first half of the

academic year?

2a. What personal background factors influence the language contact of the participants?

2b. How did the COVID-19 lockdown in the UK in Spring 2020 affect participants' informal language contact and SA experience?

3. What linguistic factors (corpus frequency, semantic opacity), personal attributes (overall English proficiency, L1) and L2 exposure (language contact, social networks) influence PV development in these participants during study abroad?

1.3 Rationale and significance of the study

There are several rationales for conducting the current study. Regarding participants, there are three common models of SA, namely, the US student model, the European student model, and the international student model (Isabelli-García & Isabelli, 2020). The US model aims to cultivate well-rounded persons by sending students abroad through university-organized SA programs to learn about different cultures. A multitude of high-quality studies were carried out on American students by Dewey and his colleagues (Dewey et al., 2014). The European model, commonly called 'residence abroad', is characterized by the popular ERASMUS program which enables students to study and live in other EU countries, paying the same fee as domestic students. For instance, the overseas L2 learning experiences of British students and Spanish students were documented in the influential LANGSNAP project (Mitchell et al., 2017) and the SALA project (Pérez-Vidal, 2014) respectively.

In the international student model, individuals enroll directly in the host university's degree programs independently and some attend pre-sessional language courses to prepare themselves for the degree study (Isabelli-García & Isabelli, 2020). The time and financial investment from this group of learners are considerable, so their stakes may be higher than students in the former two models. Traditionally, the US cohort has

been the focus of the SA literature, followed by increasing attention paid to the European cohort (Baten, 2020). Research on non-Caucasian students is rather sparse and has only sprung up in recent years (Hirakawa et al., 2018; Jackson, 2017, 2018; Khorshidi, 2013; Kraut, 2017; Rasouli Khorshidi, 2013; Schwieter et al., 2021; Tavakoli, 2018), warranting greater research attention.

As for program duration, there are rough categorizations of short-term (less than eight weeks), mid-length (one semester or over 16 weeks), and long-term SA programs (one academic year or 32 weeks and above). As most US or European students take part in short-term summer schools or semester-long SA programs, longitudinal studies on academic-year sojourns are somewhat limited (Isabelli-García & Isabelli, 2020). Thus, the current study addresses the scarcity of research in this regard by conducting a year-long study on international students from developing countries studying abroad for a degree.

The significance of the current study will be outlined according to each of the research questions. For RQ1, the selection of PVs as the dependent variable was both a personal choice and an academic choice. On one hand, the researcher, as an English teacher and an international student with rich SA experience, was often amazed by a plethora of PV usage in daily life and puzzled by some PVs she encountered in mass media or email correspondence. On the other hand, with growing attention paid to formulaic language research, research on PVs has proliferated in recent years, and more concrete gaps have emerged. To deepen our understanding of PV development during SA, the current study combines a learner corpus and standardized tests to assess the PV command of learners and tracks PV development of the same cohort of learners longitudinally.

RQ2 deepens our understanding of the language contact and social networks of foundation students. A foundation program in the British context is a preparation course to help international students meet the academic and/or language requirements for degree studies at British universities. On one hand, this is a growing student population. According to Finlayson (2019), between 2012-13 and 2017-18, the number of new

foundation year students in the UK saw a sharp rise from 10,430 to 30,030. On the other hand, this is a learner cohort in the transition period of adapting to Western culture and education. Unlike students in short-term SA programs who can only touch upon the surface of local life during a short stay, this population had a whole year abroad to prepare for their formal degree study, which they also aimed to do abroad. However, their integration might be hindered by their limited L2 proficiency. This study explores the opportunities and constraints in their informal language contact in this period, during which a lot of individual differences come into play and differentiate the social interaction of the participants.

As for RQ3, there is a dearth of studies probing into the relationship between social networks and PVs. So far, studies using social networks to explain L2 development point to a generally positive relationship with large individual variations in oral proficiency (Baker-Smemoe et al., 2014; Dewey, Belnap, et al., 2013; Xie & Liu, 2021), sociolinguistic variations (Gautier & Chevrot, 2015; Kennedy Terry, 2017), pragmatic choices (C. Li et al., 2021), lexical complexity (McManus, 2019) and self-perceived writing proficiency (Strobl & Baten, 2021). This study will further explore the role of social networks in the development of formulaic language during SA.

1.4 Defining key concepts

1.4.1 Formulaic language

Formulaic language is an area of lexical research. Formulaic language is used in the current study as an umbrella term to describe the language with some sort of formulaicity, encompassing fixed idioms and routines as well as the less fixed collocations and phrases (Smith, 2014). A plethora of terms have been present in the literature. For example, Wray (2002) identified over 50 terms used to describe aspects of formulaicity, such as *chunks*, *collocations*, *fixed expressions*, *multi-word units*, *phrasal expressions*, and *prefabricated routines*, just to name a few. Differences in

historical perspectives and theoretical frameworks to a large extent lead to this “terminological medley” (Smith, 2014, p. 14). In the current study, formulaic language and formulaic sequences are not used interchangeably, with the latter referring to individual instances of formulaic language (Siyanova-Chanturia & Pellicer-Sánchez, 2018).

1.4.2 Phrasal verbs

PVs are a type of formulaic language, which are defined as “two-part verbs consisting of a lexical verb followed by a contiguous (adjacent) or noncontiguous adverbial particle” (Gardner & Davies, 2007, p. 341). Research suggests that PVs are particularly difficult for L2 learners and a number of factors might influence their test performance (Garnier & Schmitt, 2016). The SA context tends to favor L2 competencies related to social interactions (Kinging, 2011). Thus, it is worthwhile to investigate the developmental trajectory of PV during SA, considering its ubiquity in colloquial English.

1.4.3 Study-abroad experience

The phenomenon of SA has been approached by scholars from various disciplines, such as higher education, social psychology, and intercultural communication. In applied linguistics, SA as a unique context of L2 learning and an independent subfield of second language acquisition (SLA) was underpinned by seminal work by Freed (1995). From the 2000s, the dependent variables being examined in SA literature became more diverse, ranging from linguistic outcomes to non-linguistic ones (e.g., learner beliefs, motivation, identity). However, more research on SA has not led to more consistent findings in the field (Borràs & Llanes, 2019). Therefore, more carefully designed research is needed to understand the intricate interplay between the learning environment, learner traits, and learning outcomes (Magnan & Lafford, 2011).

1.4.4 Language contact

In the current study, language contact concerns the “amount of time using the L2 and how that time is used” during SA (Dewey, 2017, p. 49). The role of language contact has been widely researched in the SA literature because it is what fundamentally differentiates the domestic at-home learning context and the English-speaking SA learning context. Some research shows a relationship between types/quantity of L2 language contact and linguistic gains, whereas some research does not (Briggs Baffoe-Djan & Zhou, 2020). Further research regarding language contact is necessary given the complexity and intricacy of the L2 usage patterns among different SA cohorts.

1.4.5 Social networks

A social network is “a set of socially relevant nodes connected by one or more relations” (Barin & Wellman, 2011, p. 11). Social network analysis is an influential area of research embedded in sociology and applied in fields such as economics, psychology, and political science (Scott & Carrington, 2011). Since its ‘take-off’ in the 1970s, some linguists have started to employ this approach to probe how interpersonal relationships affect language use (Milroy & Milroy, 1985). In the SA literature, Dewey and his colleagues and researchers of the LANGSNAP project have both incorporated language contact and social network analysis in their projects on Anglophone students. Following their practice, the present study can expand our knowledge of the role of these two constructs on the L2 development of non-Anglophone SA students.

1.5 Theoretical frameworks

The research design of the present study was mainly guided by usage-based approach and cognitive theories. As McCormick (2018) argued, SA research is largely built on the notion of usage-based approach that more L2 use leads to greater gains. Thus, usage-based approach lays a theoretical foundation for the present study. Usage-based approaches view language learning primarily based on learners’ exposure to their L2 in

use (Ellis & Wulff, 2015) and they learn linguistic constructions while engaging in communication (Bybee, 2010). It is believed that learners are ‘intuitive statisticians’ (Ellis, 2006) and the frequency effect holds that they tend to learn high-frequency structures earlier than low-frequency ones (Ellis, 2002). As for the acquisition mechanism, learners employ associative learning and cognitive processing in L2 input (Ellis & Wulff, 2015). Associative learning means learning the association between form and meaning or function of a construction, while cognitive processing implies having a mental model of their language that is custom-fit to their linguistic experience at any given time (Ellis & Wulff, 2015). The learning of new phrasal construction is exemplar-based: learners associate it with the prototypes they have seen and gradually link the form, meaning, and function of the target constructions together.

In line with usage-based approach, cognitivism in SLA took inspiration from Piaget’s cognitive determinism (Maftoon & Shakouri, 2018) argues that L2 learning takes place ‘in the mind’ rather than ‘in the world’. The minds of learners systematically process information, resulting in meaningful storage and personalized mental knowledge representation (Briggs, 2014). A few influential cognitive hypotheses are particularly relevant to the SA context. For example, the input hypothesis (Krashen, 1982) contends that the more learners are exposed to comprehensible input, the more advanced they can become. Long (1983) proposes the interaction hypothesis in which the negotiation of meanings is crucial to L2 development. The output hypothesis (Swain, 1985) highlights the pivotal role of pushed output in L2 learning. In an ideal scenario, SA is a learning environment in which L2 input permeates, L2 interaction is encouraged, and L2 output is necessary. Thus, an immersive SA experience is likely to help learners reach a more advanced L2 level.

In recent decades, sociocultural theories (Lantolf & Thorne, 2007) have been proposed to challenge cognitive views of SLA. In line with this social turn, a new socio-cognitive approach was put forward by Atkinson (2002) which encompasses the interaction between mind and context. As recognized by many researchers, the influence of the mind and the influence of the world are not dichotomous. This approach bridges the

gap between cognition and social action, because, after all, “no language acquisition can be properly understood without taking into account their fundamental integration into a socially-mediated world” (Atkinson, 2002, p. 534). The present study recognizes the important role of human interaction in L2 acquisition by investigating the role of social networks as an independent variable, although it does not theoretically adhere to the sociocultural perspective.

1.6 Thesis outline

This thesis consists of eight chapters. Chapter 1 is the introductory chapter. It outlines the global landscape of SA research, identifies some gaps in the literature, introduces the scope of the study with research questions, and briefly explicates the key concepts and the underpinning theoretical framework of this study. Then, Chapter 2, the literature review, discusses existing research in four sections, being an overview of SA research, formulaic language research with a focus on PV, research on language contact and social networks, and empirical studies on the relationship between formulaic language and language contact. In particular, the pros and cons of existing instruments for measuring language contact and social networks are demonstrated. By pointing out gaps in the literature, this chapter underlines the original contributions of the current study.

Chapter 3, the methodology chapter, delineates the research design of the current study. In terms of the sample, the features of the foundation programs and the participants’ characteristics are introduced. The instruments (i.e., the PV tests, the language contact questionnaires, and supplementary qualitative instruments) are presented. The procedures for each round of data collection are shown in flow charts. The data analysis methods are presented in the order of research questions. Finally, the methodological limitations and ethical considerations are discussed with transparency.

The findings are then presented in three chapters (Chapters 4–6), each devoted to one of the three main RQs, respectively: PV development, language contact development,

and the relationship between them. Chapter 4 is divided into PV test data and learner corpus data and each thread of data is analyzed longitudinally from Time 1 to Time 3. Chapter 5 is structured by pre-COVID data (Time 1 and Time 2) and post-COVID data (Time 3), because language contact patterns before and after the COVID-19 outbreak are inherently different, given all the lockdown measures. In each period, quantitative data and qualitative data are presented. Chapter 6 focuses on the relationship between the independent variables and dependent variables, and this chapter is organized by the types of analyses, from quantitative findings to qualitative findings.

Chapter 7, the discussion chapter, interprets the findings by RQs. The researcher refers to the literature on formulaic language and language contact to explain the results. In particular, based on the post-COVID social network findings, social support theory, which is not covered in the literature review, is also drawn on. Lastly, Chapter 8, a conclusion chapter, highlights the significance of this study, proposes implications, and provides recommendations for future research.

CHAPTER II LITERATURE REVIEW

This literature review is structured according to the key constructs of the present study. Firstly, the core issues of study abroad (SA) research will be reviewed to provide a backdrop for the present project. Second, the current state of formulaic language research and phrasal verb (PV) research will be provided to demonstrate how it is acquired and what factors influence its acquisition. Third, the instruments to measure informal language contact and social networks of SA participants will be reviewed with findings from empirical studies. Next, relevant studies on the relationship between formulaic language development and language contact in the SA context will be discussed. Finally, the original contribution of the current study to academia will be highlighted.

2.1 Study abroad research

SA is broadly defined as “an academic experience that allows students to complete part of their degree program through educational activities outside their country” (Sanz & Morales-Front, 2018, p. 1). Seminal SA research can be dated back to Carroll (1967), but it was not until the 1990s that Freed (1995) pioneered the field by synthesizing studies conducted in the SA context and first framing SA research within the broad spectrum of SLA (Collentine, 2009). In the first stage of SA research, overall L2 abilities (i.e., proficiency) were examined by pre-tests and post-tests, with predominantly quantitative measures. In line with the social turn of SLA (Block, 2003), the second stage of SA research started to pay more attention to the non-linguistic domains (i.e., learners’ individual differences and their living experiences) with qualitative measures (Magnan & Lafford, 2011). It should be noted that Magnan and Lafford (2011) summarized these stages just to roughly show the advancement and direction of progression of SA research and there is no clear cut-off between these periods. Currently, SA research is moving toward a multidimensional perspective and

a more comprehensive research design, which aims to link contextual or individual learner variables with linguistic gains (Grey et al., 2015; Leonard & Shea, 2017). In other words, researchers are trying to investigate multiple variables at the same time, to explain the complex and dynamic interplay between the learning environment, learner traits, and learning outcomes, employing mixed-methods designs (Magnan & Lafford, 2011). Since the current study chooses a type of vocabulary as the dependent variable, empirical studies on vocabulary will be presented in more detail, after first outlining the findings on other linguistic and non-linguistic outcomes.

2.1.1 Linguistic and non-linguistic outcomes

In SLA, a great deal of research has been published regarding linguistic gains during SA and synthesized by several review articles (Borràs & Llanes, 2019; Isabelli-García et al., 2018). The most well-researched area in SA research is oral proficiency of L2 learners (Llanes, 2011). Evidence suggests that SA facilitates the oral fluency development of students, be it compared with students studying domestically at-home, or compared with their own performance over time (Freed, Segalowitz, et al., 2004; Pérez-Vidal, 2014; Segalowitz & Freed, 2004). The Oral Proficiency Interview (OPI) (American Council on the Teaching of Foreign languages, 1987) was widely adopted to collect speech data from participants, and from these data speaking proficiency was often measured using complexity, accuracy, and fluency. Segalowitz and Freed (2004) found that American learners studying Spanish abroad showed greater gains in fluency than at-home learners. Freed, Segalowitz, et al. (2004) compared American students studying French in SA, at-home, and domestic immersion contexts. The results suggested that the SA group outperformed the other two groups in terms of speech fluidity, but the domestic immersion group made more significant gains than the SA group in various measures. The authors inferred that the paucity of L2 use by students in the SA context might be the main reason.

Positive influences of SA have also been observed in areas such as listening, pragmatics, and pronunciation. In terms of listening, American students (Cubillos et al., 2008) and

Spanish students (Llanes & Muñoz, 2009) abroad saw positive changes. Regarding pragmatics, American students in China made significant gains in Chinese speech act production (N. Taguchi et al., 2016) and American students studying Spanish abroad saw improvement in the pragmatic appropriateness of their requests and apologies (Shively & Cohen, 2008). As for pronunciation, most studies presented positive outcomes after SA (Díaz-Campos, 2004, 2006). Díaz-Campos noted that the SA participants outperformed their at-home counterparts in conversational tasks (2006) but not in reading tasks (2004), suggesting that the benefit of SA might vary depending on the linguistic area of emphasis.

Conflicting results on the outcomes of SA were seen in reading, writing, and grammar. Regarding reading competence, Dewey (2004) did not find significant improvement in the SA group except for higher confidence, while Kraut (2017) revealed gains in reading after just a short stay. As for writing development, there is evidence that writing improvements in SA, if any, appear later and less obviously (Sasaki, 2011; Serrano et al., 2012, 2014). Grammar gains were also barely noticeable in SA. Collentine (2004) and Hirakawa et al. (2018) even found that the at-home group outperformed the SA group, possibly because explicit formal instruction placed an emphasis on those areas.

The non-linguistic outcomes of SA in previous research range from identity, personality, learner beliefs to motivation, just to name a few. Some studies have highlighted identity development as a result of SA. Studies under this theme tend to collect a multitude of qualitative data from interviews, narratives, journals, and diaries to reveal the subtle changes in learner identity (Benson et al., 2012; Maeder-Qian, 2017; J. A. Stewart, 2010). When it comes to personality, Tracy-Ventura et al. (2016) found that there was a significant change in the emotional stability factor of British students after SA, but not in other aspects of personality. Regarding learner beliefs, Amuzie and Winke (2009) revealed that the longer the stay was, the more awareness students had regarding learner autonomy and the less importance they attached to assistance from teachers, which pointed toward an overall growth in learner independence. As for L2 learning motivation, Allen (2010) found that SA only enhanced L2 motivation for linguistically

motivated students but not for career-oriented students. Further, Allen and Herron (2003) also found that L2 integrative motivation remained unchanged after SA, but L2 anxiety decreased significantly.

Similar to other areas of SLA, arriving at consensus on the advantages of SA is extremely difficult, given the variations of SA experiences, such as student proficiency levels, program length, SA destinations, control group use, data elicitation measures, and analyses, among other issues (Isabelli-García & Isabelli, 2020). Further, even within the same study, clear-cut results were not always available, as the benefits of SA were only found on some, but not all the parameters of the target construct. Overall, SA exerts a positive influence on L2 development, but there are many conflicting results, suggesting that the advantage of SA experiences should not be overgeneralized.

2.1.2 Vocabulary development and language contact

It is clear that vocabulary knowledge grows significantly during SA (Borràs & Llanes, 2021; Briggs, 2015a; Dewaele & Regan, 2001; Dewey, 2008; Fitzpatrick, 2012; Foster, 2009; Jiménez, 2010; Milton & Meara, 1995; Pizziconi, 2017; Tracy-Ventura, 2017), though its advantage compared with the intensive domestic environment is less clear (Collentine, 2004; Segalowitz & Freed, 2004). For example, SA participants witnessed significant development in lexical diversity (Foster, 2009), lexical sophistication (Tracy-Ventura, 2017), lexical organization (Ife et al., 2000; Jiménez, 2010), lexical access (Sunderman & Kroll, 2009), lexical complexity (Serrano et al., 2011), and produced fewer lexical errors (Llanes & Muñoz, 2009).

Two studies researching vocabulary size did not identify positive relationships with language contact. Milton and Meara (1995) saw a significant lexical improvement in 53 European ERASMUS exchange students in the UK over a six-month exchange period: $t(52) = 9.12, p < .001$, and an estimated annual growth rate of 2,600 words per year. The instrument was the Eurocenters Vocabulary Size Tests (EVST), a computerized Yes/No test format with a ceiling of 10,000 words, with the caveat common to multiple-choice questions — the guessing effect (J. Stewart & White, 2011).

Contrary to assumptions about language gains, this study revealed a weak but significant negative relationship between hours of study and vocabulary gains. It was speculated that less outgoing students might spend more time in the familiar atmosphere of the lecture hall and the library, so time spent on formal study was an indication of their low integration level, rather than a causal factor of L2 vocabulary change (Milton & Meara, 1995). Accommodation types and friendship patterns (the number of English-L1 friends and English-speaking neighbors) did not have an influence on vocabulary gain. Briggs (2015a) explored the receptive and productive vocabulary development of 241 mixed-L1 private language school students studying in the UK for 6 to 20 weeks. Although the sample made significant gains in both the receptive test ($\eta^2 = .014$) and the productive test ($\eta^2 = .016$), language contact was not significantly related to these gains in either correlation analysis or multiple regression. It was inferred that this sample might not have sufficient out-of-class contact with English or enough contact that was helpful in terms of acquiring new L2 lexis.

Nevertheless, two studies measuring specific aspects of vocabulary knowledge saw the influences of language contact. In Serrano et al. (2012), the sample was 14 Spanish university students studying in the UK for one academic year. The same elicitation measures for oral narratives, written tasks, and self-reported questionnaires were administered three times (September, December, and May). Results showed that oral lexical richness growth reached significance from Time 1 to Time 2 ($Z = -2.48, p = .013$, Cohen's $d = 0.54$), but this was not the case for written lexical richness at other time points (i.e., Time 1 to Time 2, Time 1 to Time 3, Time 2 to Time 3). Questionnaire results showed that the students who spent less time with L1-speaking peers improved their lexical richness more than those who spent more time with L1 friends (written lexical richness: $Z = -2.34, p = .02$, Cohen's $d = 0.62$). A limitation of this study is the uneven intervals between the three data collection points, namely, three months and five months. Thus, this study might not be able to objectively reflect the development rate of lexical richness.

In between-group comparisons, Dewey (2008) conducted a comprehensive

investigation of the vocabulary development of 56 American students learning Japanese in a SA context, an intensive domestic immersion context, and a domestic formal classroom context. Three tests were employed: a Vocabulary Matching Test to measure receptive knowledge, a Vocabulary Knowledge Scale to assess productive knowledge, and a Situational Vocabulary Test to assess learners' ability to appropriately choose vocabulary according to everyday situations. It seems that the third test favored SA participants by nature, but the author argued that assessment used in SA research "ought to focus on areas where benefits are likely in order to determine whether those benefits do exist" (Dewey, 2008, p. 132). The result showed that it was only in the Situational Vocabulary Test that the SA group significantly outperformed both groups: $F(2,52) = 8.91, p < .001$. In the other two tests, both the SA and the domestic immersion groups outperformed the formal classroom group [$F(2,52) = 11.89, p < .001$] and there was no significant difference between the SA and the domestic immersion group ($p > .05$). Overall, SA was beneficial to vocabulary development, although gains were similar between the SA and the domestic immersion settings.

Through a questionnaire, learner diaries, and interviews, Dewey (2008) found that interactive contact played a key role in vocabulary gains. For the SA group, the strongest correlation was found between time spent speaking with Japanese friends and the Vocabulary Knowledge Scale ($r = .459, p < .05$) and Situational Vocabulary Test results ($r = .610, p < .05$). Interestingly, time spent reading emails or web pages in Japanese was negatively correlated to the Vocabulary Knowledge Scale scores ($r = -.468, p < .05$). For the immersion group, time spent speaking with non-native speakers was significantly correlated with Vocabulary Knowledge Scale results ($r = .611, p < .05$). Serrano et al. (2012) and Dewey (2008) highlighted the importance of interaction in the advancement of L2 vocabulary. According to the output hypothesis (Swain, 1985), L2 learners are likely to make progress when they notice a gap in knowledge in their pushed output, which is the output that "reflects what learners can produce when they are pushed to use target language accurately and concisely" (R. Ellis, 2003, p. 349).

Finally, a case study by Fitzpatrick (2012) shed light on the incremental vocabulary

development during SA. Lexical knowledge was elicited through a word association task (Lex-30, Meara & Fitzpatrick, 2000) completed by one learner six times a year. This study revealed a gradual increase in some aspects of vocabulary knowledge (the number of collocations and native-like associations), but striking inconsistencies in others (e.g., word form, form-meaning connections, orthography), exposing the complex and multidimensional nature of lexical acquisition. Based on the above discussion, it is clear that SA contributes to lexical development, but the inconsistent findings between language contact and vocabulary gains imply that the learners might not fully take advantage of L2 language contact when they study abroad or their L2 contact was not fully captured by the instruments, which will be discussed in Section 2.3.

2.1.3 Variability in the impact of study-abroad on SLA

Variation is a key theme in SA research in SLA and it should be expected that differences in program features and participants' attributes directly influence learning outcomes. Several factors that could differentiate learning outcomes during SA should be considered (e.g., proficiency, age, L1, motivation, living arrangement, program location). First, the role of pre-program L2 proficiency (that is, the proficiency level of a learner at the outset of their SA experience) in predicting L2 gains deserves attention. Previous research has exhibited certain contradictions regarding what level of pre-departure proficiency would lead to optimal gain (Köylü, 2021). Some studies argued that learners need to have basic L2 knowledge to benefit from SA (DeKeyser, 2014; Magnan & Lafford, 2011), so higher initial proficiency favors L2 development (Borràs & Llanes, 2021; Leonard & Shea, 2017). Nevertheless, there is also ample evidence pointing to an opposite argument that lower-level learners benefit more in L2 development than advanced learners (Llanes & Muñoz, 2009; Milton & Meara, 1995; Pizziconi, 2017; Zaytseva, 2016), especially when using instruments such as the OPT (Llanes, 2019). Further, some studies found no correlation between predeparture proficiency levels and the amount of L2 gains over a semester in an ELFSA context

(Llanes, 2019). With this complexity remaining, research has yet to ascertain whether there is any difference in the PV development trajectory between advanced learners and lower-level learners.

As for the role of age, the findings are mixed. Llanes and Muñoz (2009) found that age did not significantly differentiate the L2 gains of participants aged 13–22 in a short-term SA program. Absolute gains refer to the percentage of gains in the full scores, while relative gains refer to the percentage of gains of the obtained scores. However, Llanes and Muñoz (2013) found that adults made more absolute gains than children, while children made more relative gains than adults in writing measures in SA. Regarding L1, past studies have often been conducted on learners from the same country to control this factor. Jensen and Howard (2014) found some differences in the L2 gains of Chinese (n=8) and French (n=10) learners of English: Chinese learners produced fewer words per clause at the end of SA and had higher error rates than French learners. Although there is evidence that linguistic distance between the L1 and the L2 affects L2 proficiency (Chiswick & Miller, 2005), due to the small sample size, the authors could not draw substantive conclusions about the role of the Chinese L1 in the development of L2 speaking.

N. Taguchi (2008) suggested that learners' personality and motivation are critical factors in determining the quality and quantity of L2 interactions, which will lead to L2 gains, such as oral proficiency (Hernández, 2010). Isabelli-García (2006) noted a recursive relationship between motivation and social integration during SA, in which the success or failure of entering the social circles of L2 speakers will, in turn, enhance or weaken L2 motivation. While some students' motivation remained unchanged (Allen & Herron, 2003), Hessel (2016) found that learners' motivation reached a peak in the first three months of SA and decreased significantly afterward. As suggested by Dewey et al. (2013; 2014), connections between motivation, social relationships, language use, and language acquisition are worth exploring further.

In addition to the above individual differences, some variabilities in the SA programs

were also explored, with mixed results. In terms of living arrangements, Spanish learners who lived with only English-speaking people experienced more gains in written lexical richness than those who lived with L1-speaking people between Time 2 and Time 3 of a year-long SA sojourn, because the former learners possibly had more L2 interaction opportunities (Serrano et al., 2012). In contrast, Rivers (1998) found that homestay students were less likely to improve their listening and speaking than their dormitory-resident counterparts, because they did not form useful communicative partnership with their hosts. With regards to program location, in a comparison of six SA destinations in the world, students in Egypt had significantly more L2 contact than students in the other five countries because of a curriculum arrangement with a speaking requirement (Dewey et al., 2014). Within the same country, Briggs (2015a) did not identify a significant difference in the vocabulary growth of participants studying in London and Oxford. Lastly, regarding lengths of stay, the evidence was conflicting. While some studies identified that a longer stay was beneficial to vocabulary gains (Briggs, 2015a; Ife et al., 2000), lengths of stay did not play a role in the English oral and writing skills of some Spanish students (Llanes & Serrano, 2011) or the English conventional expressions of international students in the US (Bardovi-Harlig & Bastos, 2011). It seems that the role of length of stay may differ by the specific L2 skill being measured (Jensen & Howard, 2014), as oral skills were more likely to witness gains early (Hardison, 2014; Llanes & Muñoz, 2009).

The brief overview above introduces the core issues and the latest trends in current SA research and provides an empirical background for the current study. Faced with the myth of SA as the *sine qua non* recommendation for L2 learning and the contrasting results of previous empirical inquiries, more attention is needed to investigate the factors that play a role in SLA during the sojourn abroad. Examining the informal L2 contact of students studying abroad will provide us with new insights to transform SA from “a black box” to “a glass box” (N. Taguchi, 2018, p. 128) and better understand how to make the most of the SA experience to facilitate L2 learning. In the meantime, other individual differences (e.g., overall English proficiency, L1, and motivation) will

be explored regarding their influence on the current sample.

2.2 Researching phrasal verbs

Vocabulary research encompasses investigations of one-word vocabulary and multi-word items. Before looking at PV research, this section will start with theories related to the acquisition of general (one-word) vocabulary and formulaic language.

2.2.1 Vocabulary learning in SLA

A variety of frameworks have been proposed to define what it means to ‘know’ a word. These frameworks commonly address three features of vocabulary: breadth (size), depth, and lexical access (fluency) (Daller et al., 2007). For example, the influential taxonomy by Nation (2001) categorizes knowledge of vocabulary into three dimensions: form, meaning, and use. Vocabulary is a facet of linguistic competence that, to a large extent, predicts overall language competence (Schmitt, 2010). According to a review by Qian and Lin (2019), vocabulary size is one of the strongest discriminating factors of L2 learners’ proficiency levels. Vocabulary knowledge has been found to predict reading comprehension and listening comprehension (Cheng & Matthews, 2018; Hacking & Tschirner, 2017; S. Zhang & Zhang, 2020). There is also robust evidence supporting the positive impact of vocabulary knowledge on L2 speaking (Janebi Enayat & Derakhshan, 2021; Uchihara & Clenton, 2020) and writing (Dabbagh & Janebi Enayat, 2019; Kilic, 2019).

L2 learners acquire new vocabulary in two ways: incidentally and intentionally (Hulstijn, 2003). According to Webb (2007, p. 47), “incidental learning is a gradual process in which gains are made in small increments with repeated encounters needed to gain full knowledge of a word”, while intentional learning refers to learning activities involving conscious cognitive operation (Hulstijn, 2003) which often occur during formal instruction. Since the current study concerns out-of-class L2 contact, only incidental learning will be introduced.

Laufer and Hulstijn (2001) applied the notion of depth of processing (Craig & Lockhart, 1972; Craig & Tulving, 1975) to vocabulary acquisition and put forward the Involvement Load Hypothesis to explain incidental vocabulary learning. Depth of processing means that the chance of some pieces of new information being stored into long-term memory is not determined by the length of time that it is held in short-term memory but rather by the shallowness or depth with which it is initially processed.

Laufer and Hulstijn (2001) proposed a motivational-cognitive construct of involvement. In real-life communicative situations, the involvement load encompasses three components: the *need* to use a particular word (a motivational, noncognitive dimension), the *search* to find the meaning of this word (the attempt to find the meaning of an unknown L2 word or find the L2 word form expressing a concept), and the *evaluation* of the usage of this word (comparison of a word with other words, a meaning with other meanings). It is argued that retention of unfamiliar words is, generally, conditional upon the degree of involvement in processing these words and “the greater the involvement load, the better the retention” (Hulstijn & Laufer, 2001, p. 545). Yanagisawa and Webb (2021) found that the *evaluation* component contributed to the greatest amount of learning, followed by *need*, whereas *search* did not contribute to learning. They also identified that vocabulary test format and frequency of exposure to words moderated learning gains, and higher involvement load had a greater impact on learning than longer time on task. These findings suggest that the quality and frequency of L2 contact are important for incidental vocabulary acquisition.

2.2.2 Acquisition of formulaic language

Formulaic language has occupied a prominent role in lexical studies (Siyanova-Chanturia & Pellicer-Sánchez, 2018; Wray, 2013). Wray (2002, p. 9) defined formulaic language as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”. Formulaic language research appears in psycholinguistics, corpus

linguistics, cognitive linguistics, and language education with various subcategories identified (e.g., *multi-word units*, *idioms*, *binomials*, *collocations*, *phrasal expressions*, and *lexical bundles*). As mentioned in the introduction chapter, formulaic language and formulaic sequences are not used interchangeably in this thesis. The former is a coverall term that includes the latter. Formulaic language is estimated to constitute a significant portion of any competent speaker's spoken language. Erman and Warren (2000) calculated that formulaic sequences of various types constituted 58.6% of the spoken English discourse and 52.3% of the written discourse. Foster (2001) and Biber et al. (1999) judged that over 30% of the native speech was made up of formulaic language. Besides, formulaic language not only provides a useful purpose in communication but also helps learners to be more fluent (Conklin & Schmitt, 2012), justifying the necessity of mastering formulaic language.

Conklin and Schmitt (2008, 2012) found that, unlike English L1 speakers, there is mixed evidence regarding the processing advantage of formulaic language on English L2 speakers. While very proficient L2 speakers sometimes exhibit processing advantages similar to native speakers, less proficient learners often process formulaic language in a word-by-word manner similar to non-formulaic language (Conklin & Schmitt, 2012). Furthermore, if the formulaic language is idiomatic, the figurative meanings can be much more difficult for L2 speakers to process than literal non-formulaic language.

There are two different views on the acquisition mechanism of how formulaic language is stored, accessed, and retrieved from the mental lexicon. Ellis (2002) – in line with usage-based perspectives on SLA – resorted to the frequency effect to explain adult L2 learners acquiring collocation. In lay terms, the more often one sees/uses the construction, the more likely it is to be learned. However, Wray (2002) maintains that adult L2 learners take a non-formulaic approach in which they memorize the meaning of individual words when acquiring formulaic sequences. In an experimental setting, Durrant and Schmitt (2010) recruited a control group that was only exposed to the target items once and two experimental groups with verbal repetition or varied repetition of

the target items. This study dealt with implicit learning only, and no formal instruction was provided. Significant gains in the experimental groups demonstrated that limited collocation knowledge should be attributed to the lack of exposure rather than the single-word-based learning approach, which favors the former model.

Thus, the current study adopts the usage-based perspective to research incidental PV acquisition. Above all, the linguistic input learners receive is the primary source for their L2 acquisition (N. C. Ellis & Wulff, 2015). It is argued that PV knowledge is accumulated in the lexicon through encounters in concrete situations and the PVs are stored and processed as a whole unit by associative learning and the ‘chunking’ process (Wulff, 2018). ‘Chunking’ is “the development of permanent sets of associative connections in long-term memory” and chunks form the basis for automaticity and fluency in language use (N. C. Ellis, 1996, p. 107). For example, after automatization, students have a deeper impression of the meaning of “hold up” (delay) and can understand it as a whole unit.

It should be noted that, in usage-based learning, frequency is not the only factor influencing the effectiveness of uptake. Distribution of PV in the input, recency, salience (i.e., how much it stands out relative to its context), redundancy (i.e., redundant for the listener’s understanding for an utterance) of the PV exemplars in the input, perception of form-functional prototypicality, and contingency of form-function mappings can mediate learning (Wulff, 2018). She summarized that unevenly dispersed distribution, high recency, high salience, low redundancy, strong prototypicality of PV exemplars, and reliable form-function associations favor the acquisition of the target item. It is hard to say whether one learning context is definitely superior to another, given the interplay of all these factors in the L2 input.

2.2.3 Research on phrasal verbs

There have long been two rival approaches to defining formulaic language. The phraseology perspective underscores the opaque semantic features of formulaic language, while the frequency-based perspective focuses heavily on the statistical

probability of word co-occurrence in the criteria (Barfield & Gyllstad, 2009). With the development of corpus linguistics, and considering the potential subjective judgment in semantic criteria, the frequency approach seems to prevail (Gablasova et al., 2017). The traditional phraseology approach relies on human analysts to identify formulaic language, so only a limited number of items can be analyzed and researched via this approach, though the case-by-case accuracy might be higher. The frequency-based approach is gaining popularity because it makes use of language databases (computer corpora) to analyze language formulaicity, which is seen as more efficient, representative, and systematic.

As a subcategory of formulaic language, the definition of PV is as fuzzy as the field itself. Quirk et al. (1985) argued that PVs had a syntactic criterion of “verb + particle” and a semantic criterion of having meanings beyond the separate meanings of the two components. Biber et al. (1999) devoted their attention to distinguishing four types of PVs: 1) phrasal verbs (e.g. *pick up*). 2) prepositional verbs (e.g. *look at*). 3) phrasal-prepositional verbs (e.g. *get away with*). 4) other multi-word verb constructions, notably: verb + noun phrase (+ preposition) e.g. *take a look (at)*; verb + prepositional phrase, e.g. *take into account*. Gardner and Davies (2007) believed that ambiguity and inconsistency will eventually lead to frustrations for English teachers and students, so they proposed a functional and objective definition that PVs are “two-part verbs consisting of a lexical verb followed by a contiguous (adjacent) or noncontiguous adverbial particle” (Gardner & Davies, 2007, p. 341) without the semantic criterion. This is the definition that is adopted by the current study because it largely avoids the ambiguity of the semantic criterion in some other PV definitions and this definition has been widely adopted by several key studies in the field (e.g., Garnier & Schmitt, 2015; Liu, 2011; Liu & Myers, 2018).

The rationale for choosing PV as a dependent variable of interest in SA research is based on its importance and its difficulty. First, the more often a word is used in a language, the more important it is to be learned (N. C. Ellis, 2002). Gardner and Davies (2007) found that, on average, one PV would appear in every 150 words of English in

the British National Corpus (BNC). Furthermore, they are crucial to fluent English and native-like spoken proficiency. PVs are widely used in the spoken informal discourse of L1 English speakers (Garnier & Schmitt, 2015). Therefore, the ubiquity of PVs necessitates research. Previous research showed that L2 English learners tend to overuse single words and underuse PVs compared with L1 English speakers, making their language unnatural or non-idiomatic (Siyanova & Schmitt, 2007). Therefore, the mastery of PVs can be regarded as one of the hallmarks of advanced English competencies, which is useful for living and studying in an Anglophone country.

The difficulty of PVs stems from a few aspects. First, PVs are highly polysemous. Gardner and Davies (2007) found that each of the most frequent English PVs had 5.6 meanings on average. Inadequate knowledge of multiple meanings might cause confusion to learners when they encounter PVs. Second, the meanings of PV have different semantic transparency: PVs have prototypical (literal) meanings and metaphorical (figurative) meanings (Garnier & Schmitt, 2015). Third, PVs have the syntactic peculiarity of particle movement (Gardner & Davies, 2018) (e.g. *let you down* VS **let down you*). The separable PVs might lead to lower salience for noticing because learners might try to decode their meaning from the individual words rather than as a whole unit (Garnier & Schmitt, 2015). Moreover, PVs often contain delexicalized verbs (e.g., *make, get, take*), making their meanings opaque and hard to decipher (Schmitt & Redwood, 2011). Lastly, learners whose L1 lacks such a structure might be unfamiliar with this construction, which increases the difficulty of mastering PVs (Darwin & Gray, 1999; Liao & Fukuya, 2004). Based on the above points, since PVs are a “notoriously challenging aspect of the English language” (Gardner & Davies, 2007, p. 339), more research is needed to understand the factors that facilitate or inhibit PV acquisition.

Existing research on PVs can be summarized into three strands. The first line is identifying the PV items worth learning by corpus analysis of native-speaker corpora (Gardner & Davies, 2007, 2018; Liu, 2011; Liu & Myers, 2018). The second strand is measuring L2 learners’ knowledge by pen-and-paper testing or corpus analysis (Garnier & Schmitt, 2016; Kamarudin et al., 2019; Schmitt & Redwood, 2011; Waibel, 2007;

Zarifi & Mukundan, 2014; X. Zhang & Wen, 2019). The third theme is improving learning outcomes through intervention studies with an experimental design in the classroom (Lu & Sun, 2017; Strong & Boers, 2019a, 2019b; Teng, 2020; White, 2012). What the present review focuses on is the second strand: the measurement of learners' PV knowledge. Before elaborating on relevant studies, the PV lists in the first strand will be briefly introduced.

Admittedly, PVs are constantly evolving, with new PVs created every day, and nuanced meanings of PVs might vary according to dialect (Bolinger, 1971). In recent decades, three high-frequency PV lists were compiled based on different native speakers' mega-corpora to inform teaching and learning (Gardner & Davies, 2007; Garnier & Schmitt, 2015; Liu, 2011). The list by Gardner and Davies (2007) contains a total of 100 PVs made up of the top 20 PV-producing lexical verbs (e.g., *come*, *go*, *get*, and *take*) based on the BNC. Taking this one step further, Liu (2011) compiled a PV list with 150 high-frequency PVs from the Corpus of Contemporary American English (COCA). This list includes the 100 PVs from Gardner and Davies (2007) and places no restriction on the frequency of lexical verbs. This comprehensive list also provides the PV ranks and frequencies in the BNC, which demonstrates similarities in the high-frequency PVs in both corpora. Garnier and Schmitt (2015) analyzed the polysemy phenomenon on every PV from Liu (2011) and turned it into a pedagogical list called the PHaVE List, with their most frequent meanings presented. The current study follows the PV corpus frequency in BNC and COCA in Liu (2011) and tests their most frequent meanings from the PHaVE List (Garnier & Schmitt, 2015). The PHaVE list is an elaboration and extension of the PV list by Liu (2011). Therefore, using this list in the current study drew on the latest and more comprehensive high-frequency PV list in PV research to test students' PV knowledge. This list follows the corpus frequency from the BNC and the COCA, because these two corpora represent two most influential English variants and the English use in the majority of international mass media.

2.2.4 The assessment of phrasal verb knowledge

2.2.4a Researcher-developed tests

Two previous studies with standardized tests are key references for the current study. Schmitt and Redwood (2011) administered a blank-filling productive test followed by a multiple-choice receptive test for 60 target PVs. Among them, 50 high-frequency PVs were selected from Gardner and Davies (2007) and ten off-list PVs were chosen from student coursebooks (e.g., *pin down*, *take after*). For the 50 in-list PVs, they chose the ones that students presumably had the most exposure to (e.g., *sit down*, *get up*) without a systematic approach. The tests were administered to 68 intermediate and upper-intermediate mixed-L1 learners at three private language schools in the UK. In the productive and receptive tests, learners recognized about 48% and 65% of the target PVs, respectively. This result supported the conceptualization that receptive knowledge precedes productive knowledge (Henriksen, 1999). Two measures in the receptive test enhanced validity. As shown in Figure 2.1, they made sure all the choices contained the same word (e.g. *set*), because the initials were already shown in the productive test (s ____). They also included option *E* (*I don't know*), to minimize random guessing. However, a limitation is that, although the test was piloted three times, the authors did not present the Cronbach's α of the items or other validation evidence in the publication.

Figure 2.1 Productive test and receptive test in Schmitt and Redwood (2011)

The police s_____ u_____ roadblocks to stop people driving into the city centre. (<i>build, erect</i>)				
The police _____ _____ roadblocks to stop people driving into the city centre. (<i>build, erect</i>)				
A. set in	B. set up	C. set on	D. set at	E.?

Garnier and Schmitt (2016) assessed the productive knowledge of 40 common polysemous PVs in the PHaVe list (Garnier & Schmitt, 2015) of 128 English-major

university students in Chile, with 100 cloze-format questions (testing 2–4 meanings for each PV). The target PVs were selected by a random number generator to ensure systematicity, but as a result, some selected PVs had low salience in daily life (e.g., *look back*, *give out*) with neither highest frequency nor high idiomaticity. This test has a high internal consistency (Cronbach’s $\alpha = .95$) but it was not properly validated (Garnier, 2016). Testing polysemy is desirable in the field, but it also brings two challenges. First, although the instruction specified that “some of the sentences use the same phrasal verb” (Garnier & Schmitt, 2016, p. 40), participants did not know some PVs would be tested up to four times, which might pose difficulty to them. Second, some PV meanings usually appear in certain contexts (e.g., politics, criminology, science), so the prompts were relatively difficult with many technical terms. For example, the fourth meaning of *break down* is “undergo chemical decomposition; separate into different substances”, so the corresponding test item is: “*Digestion b___ d___ substances into small molecules*”. (*decomposes*) (p.40). If the participants could not answer items like this, it remains uncertain whether they did not know the target PV or did not understand the prompt.

The results showed that the participants on average obtained 40.6% of the total scores. It was found that high-frequency PVs were recalled better than low-frequency PVs, but students’ command of specific meanings did not drop in parallel to their coverage rankings. As Garnier and Schmitt (2016, p. 34) concluded, “for most PVs, the percentage known bounces up and down among the coverage ranks with no clear pattern”. Overall, these findings lent support to the frequency effect (N. C. Ellis, 2002) and demonstrated students’ limited command of polysemous PVs. Thus, considering the English proficiency of the current sample, polysemy was not tested.

X. Zhang and Wen (2019) assessed Chinese university students’ receptive knowledge of polysemous PVs via a 160-item meaning acceptability judgment test. They selected 50 PVs with two meanings from the PHaVe List and included 60 filler items to avoid priming effects. Different from typical form-meaning mapping, the participants were asked to judge the use of PVs in the sentences on a five-point scale from ‘totally

unacceptable' to 'completely acceptable'. They also investigated a variety of potential predictor variables, such as cognitive factors (entrenchment and preemption), linguistic factors (corpus frequency and semantic transparency), and learner factors (L2 proficiency and L2 exposure), through mixed-effects modeling.

Some other tasks examined the avoidance of PV in L2 learners. For example, Chinese learners were found to avoid PVs, measured by multiple-choice questions, translation, and story-retell tasks (Becker, 2014; Liao & Fukuya, 2004). Dagut and Laufer (1985) documented the avoidance phenomenon among Hebrew-speaking students in translation and multiple-choice tests. They also found that literal meanings of PVs were used more frequently than figurative meanings. Avoidance of PVs was also found in Egyptian learners (El-Dakhs, 2016) and Jordanian students (Haider et al., 2020). Haider et al. (2020) found that advanced learners used more PVs than one-word lexical items compared with the intermediate groups. In short, L2 learners tend to refrain from using PVs in written production, and it remains a question whether this trend may change or persist in speech samples if they study abroad.

2.2.4b Corpus analysis

Corpus analysis was widely adopted to measure the free PV production of L2 learners. Waibel (2007) quantitatively compared written PV use by German and Italian learners in the International Corpus of Learner English (ICLE) and qualitatively analyzed the unnaturalness, creativity, style, and expressivity of their PV use. Ryoo (2013) analyzed PV production from 257 essays written by Korean university students in English essay contests and compared them against the BNC. She found that the most frequent PV combinations used by native speakers were recognized and utilized by the Korean students, while less frequent combinations in the BNC rarely appeared in the learner corpus. In Malaysia, Zafari and Mukundan (2014) analyzed the creativity and unnaturalness of PV production in the English of Malaysian Students (EMAS) corpus. They found that learners often produced unacceptable PV combinations and made use of some common PVs with meanings not attested by any dictionaries, showing

deviations from natural use. In China, Wei (2021) analyzed PV use in the Spoken English Corpus of Chinese Learners (SECCL). Results showed that the progress of oral PV use was evident from intermediate students to advanced students (measured by their oral test performance), and PV, in general, occurred less frequently in speaking compared with the frequency of their single-word synonyms. Based on extant corpus research, in the current study, frequency, accuracy, and sophistication could be feasible directions for analyzing PV use quantitatively in learner corpora.

2.2.4c Combining test data and learner corpora

Tracy-Ventura (2017) pioneered methods that combined corpora and experimental data to assess the knowledge and use of sophisticated vocabulary by English L2 learners. In terms of PVs, Kamarudin et al. (2019) also adopted a triangulation method. They assessed the PV knowledge of 480 Malaysian secondary school students using a 40-item receptive test adapted from Siyanova and Schmitt (2007) with a Cronbach's α of .806. Further, they conducted a qualitative analysis of five low-scoring PVs (i.e., *put up*, *come across*, *give up*, *run into*, *come out*) from the receptive test in the EMAS corpus which includes written and verbal production. The researchers adapted the test prompts into a teenager-friendly dialogue format in daily life contexts, with an unsystematic design of distractors. For example, in the item in Figure 2.2, none of the choices used the same lexical verb and the wrong choices had very different meanings from the correct answer. In some other items, two PVs with the same lexical verbs appeared in the choices (e.g., *put up*, *put on*), providing extra hints to the learners. The participants obtained a mean score of 75%, and 36.5% of the participants scored over 80% of the full score, signaling an evident *ceiling effect* based on the histogram. Qualitative analysis of the learner corpus revealed unfamiliarity with many common meanings of PVs and nonstandard use of PVs by combining meanings of each element in the PV construction.

Figure 2.2 Receptive PV test in Kamarudin et al. (2019)

A: "Many people are dying of lung cancer nowadays."
B: "Yeah. Smoking could be one of the reasons, I guess."
A: "I hope my dad will _____ smoking."
A. put down B. point out
C. give up D. throw away

Based on the above review, two gaps in the literature are addressed in the current study. First, a receptive test, a productive test, and a learner corpus will be employed to comprehensively evaluate the PV command of L2 learners in terms of meaning, form, and use. Second, previous longitudinal PV research is limited and was usually conducted in experimental settings. The current study will track the longitudinal development of PV in a naturalistic setting for one academic year, which pushes new boundaries in the literature.

2.2.5 Factors influencing phrasal verb knowledge

2.2.5a Linguistic factors

A multitude of studies have explored how PV knowledge is mediated by different linguistic and personal factors. In terms of linguistic factors, the importance of corpus frequency is undisputable (Garnier & Schmitt, 2016; Schmitt & Redwood, 2011; X. Zhang & Wen, 2019). Vocabulary research shows that the more frequent a word, the more likely it is to be known (Schmitt, 2010), which also applies to PVs. Schmitt and Redwood (2011) revealed that corpus proficiency could predict more variance in productive PV scores (20%) than in receptive PV scores (9%). This result suggested that performance in productive tests might be more directly related to overall proficiency, and performance in receptive tests might be subject to more test design factors. It remains a question as to what extent the PV presence in native speakers' corpora resembles the actual English exposure of learners, which can partially be revealed by exploring the patterns of their L2 language contact and social networks.

Another linguistic factor that deserves attention is semantic opacity, whose role is not

entirely clear. Semantic opacity/transparency refers to the extent that the meaning of a PV could be interpreted literally. Various terms have been used in discussing the idiomaticity of PVs, such as ‘literal’, ‘transparent’, ‘non-literal’, ‘figurative’, ‘opaque’ and ‘idiomatic’. In the current study, semantic opacity and semantic transparency are used interchangeably as they fall on the opposite ends of the continuum of idiomaticity. Conklin and Schmitt (2008) identified that for high-frequency figurative formulaic sequences, there might be an intricate tension between the facilitating effects of frequency and the inhibiting effects of idiomaticity. Garnier and Schmitt (2016) found that semantic opacity was not a significant predictor of learner performance in a productive PV test, while X. Zhang and Wen (2019) found it was in a receptive PV task. Except for different test formats, the different operationalizations of this construct might lead to the different result. Garnier and Schmitt (2016) used a binary judgment of semantic opacity based on native speakers’ intuition while X. Zhang and Wen (2019) asked the participants to rate to what extent the meaning was easy to guess using a five-point Likert scale. Thus, the semantic opacity categories in X. Zhang and Wen (2019) was more fine-grained and subjective than those in Garnier and Schmitt (2016). Whether input frequency can override the influence of semantic transparency in PV acquisition in SA will be addressed by the current study.

2.2.5b Learner attributes

Regarding learner attributes, cross-sectional studies have demonstrated consistent evidence on the positive relationship between general language proficiency and PV knowledge (Kamarudin et al., 2019; Schmitt & Redwood, 2011). The relationship between single-word vocabulary and formulaic language was also positive (Gyllstad, 2005), although formulaic language competence usually lags behind single-word vocabulary (Irujo, 1993). As for whether language proficiency influences the PV development rate of learners, Schmitt, Dörnyei, et al. (2004) did not find such a correlation on formulaic sequences: that is, there is no significant correlation between formulaic language gain scores and any Vocabulary Levels Test (VLT) scores or VLT gain scores. The current study will investigate whether PV gains have a relationship

with initial English proficiency or overall English proficiency as measured by IELTS and the VLT. Third, formal classroom instruction, by and large, overlooks PV knowledge. Neither Garnier and Schmitt (2016) nor Schmitt and Redwood (2011) revealed a statistical relationship between the PV test scores of students and their years spent studying English or their weekly hours of English classes. Furthermore, extant research has revealed that L1 might exert an impact on the PV production of learners (see Section 2.2.4). Waibel (2007) found that students speaking certain L1s (e.g., French, Italian, Spanish, and Russian) underproduced PVs compared to native speakers of English; Dutch and Polish L1 learners produced a similar number of PVs as English native speakers; and German L1 speakers overproduced PVs in their writing. Thus, L1–L2 distance might play a role in PV acquisition, which favors L1 Germanic languages. The current sample does not have Germanic L1, so only the role of Chinese L1 will be examined against other non-Germanic L1s.

2.2.5c L2 exposure

Furthermore, researchers are also interested in whether engagement in L2 could influence PV knowledge. Various terms to refer to language contact exist in the SA literature, such as ‘L2 contact’, ‘L2 exposure’, ‘L2 use’, ‘language use’, ‘linguistic experience’, ‘language engagement’, ‘extramural activity’ (Briggs Baffoe-Djan & Zhou, 2020). Thus, in the current study, many of these terms will be used interchangeably. As Durrant and Schmitt (2010) argued, the weak formulaic language repertoire of English L2 learners mainly stems from a limited amount of exposure. Therefore, it is reasonable to investigate what kind of L2 learning experiences can increase effective L2 exposure for PV acquisition. Past research showed that spoken input L2 contact was conducive to higher PV command in SA (S. Zhou & Briggs Baffoe-Djan, 2022). Extensive reading, watching TV, and using social networking sites in L2 were found to significantly predict formulaic language competence, while listening to music in L2 was not (Fernández & Schmitt, 2015; Garnier & Schmitt, 2016; Schmitt & Redwood, 2011). However, some inconsistency remained between different studies and these activities were all receptive L2 activities that did not involve authentic interpersonal communication. English as a

Foreign Language (EFL) learners rarely have opportunities to speak English in daily life, but whether L2 communication can facilitate PV development among English as a Second Language (ESL) learners abroad can be examined. Section 2.3 will elaborate on measuring the L2 engagement of the sample.

2.3 Informal language contact in SA

Language contact and social networks are two distinct but highly related constructs in the SA literature because the social network patterns of L2 learners can influence the interactive language contact they receive. The present study explores these two constructs under the term ‘language contact’ because, essentially, they both tap into the quantity and quality of L2 use during SA. Measuring language contact and social networks serves the same purpose of understanding the characteristics of L2 exposure of SA participants in this study.

2.3.1 Language contact

Language contact concerns the “amount of time using the L2 and how that time is used” during SA (Dewey, 2017, p. 49). While some scholars include both in-class and out-of-class contact in their operationalization of the term (Martinsen et al., 2010), others specify that out-of-class (or ‘informal’) contact is their focus of interest (Briggs, 2016). The current study focuses on informal L2 contact because it is this feature of language contact that (should) differentiate between the domestic and SA settings and is most valued by SA learners, rather than in-class formal instruction (Briggs Baffoe-Djan & Zhou, 2020).

2.3.1.1 Language contact research in SA

Research on language contact can be categorized into two strands. One line of inquiry takes language contact as the dependent variable. Some studies attempted to document the L2 contact of SA participants as truthfully as possible (Dewey et al., 2014; Fernandez & Gates Tapia, 2016), while some recent studies documented both L1 and

L2 contact, recognizing the multilingual reality of SA experiences (García-Amaya, 2017; Martínez-Arbelaiz et al., 2017; McManus et al., 2014). It should be acknowledged that there is lack of clear-cut distinction in this research. For example, Martínez-Arbelaiz et al. (2017) found that European exchange students in Spain used three languages in their daily lives. They mostly used English in an academic setting, used their L1s on mass media and social media, and used Spanish mostly to obtain services and maintain personal relations in face-to-face conversations. While more L2 use than L1 use for SA students was sometimes reported (García-Amaya, 2017), higher L1 use might be more common for SA participants (Dewey, 2017; Trentman, 2013). Further, decreasing trajectories of L2 use were witnessed (García-Amaya, 2017), especially students' L2 speaking frequency with friends and host family members. McManus et al. (2014) found that L2 use by British students studying in France differed by scenario. L1 use was most frequent in virtual contexts and general free time, while L2 was most commonly used in the workplace/university and organized free time for small talks and service encounters. Overall, several studies show that L2 use during SA does not show an upward trend, and there is a high variance of L2 use according to tasks and contexts.

An array of studies have measured L2 contact as an independent variable to investigate the relationship between language contact and L2 development. Most of the studies in this category were conducted on oral proficiency (Baker-Smemoe et al., 2014; Freed, Segalowitz, et al., 2004; García-Amaya, 2021; Hernández, 2010; Magnan & Back, 2007; Segalowitz & Freed, 2004; Yager, 1998) and pragmatics (Shively & Cohen, 2008; N. Taguchi, 2008; N. Taguchi et al., 2016; Tang et al., 2021), with some studies on reading (Dewey, 2004), phonology (Díaz-Campos, 2004), vocabulary (see Section 2.1.2), and so on.

Among these studies, results on whether L2 gains are significantly related to language contact require clarification and delimitation. Some studies failed to detect such a relationship in reading (Dewey, 2004), oral proficiency (García-Amaya, 2021; Magnan & Back, 2007), and lexical development (Briggs, 2015a; Hardison, 2014; Pérez Vidal

& Juan-Garau, 2009) while several others did (Dewey et al., 2012; Díaz-Campos, 2004; Hernández, 2010; N. Taguchi, 2008; Yager, 1998). Although “the more contact, the higher gains” is a common conclusion, upon closer scrutiny, counterintuitive findings do exist. For example, descriptive statistics in Martinsen et al. (2010) indicated that the students with the greatest oral L2 gains spent more time speaking with roommates who were non-native L2 speakers, while those with few or no language gains spent much more time speaking with native speakers. One of the explanations offered by the authors was that maybe the participants spent more time listening to the native speakers since they were less fluent than native speakers, while students who interacted with their roommates had more equal exchanges. Magnan and Back (2007) revealed that speaking L2 with L1 peers had a negative correlation with L2 oral proficiency development (Spearman’s $\rho = -.386$, $p = .046$), implying that L2 contact might be more beneficial in authentic communication and that it matters with whom students speak L2. Further, within the same study, the complexity remained that language contact did not influence all aspects of L2 competence. For example, N. Taguchi (2008) found that the amount of speaking and reading outside class significantly correlated with the gains in pragmatics comprehension speed, but not with the accuracy of comprehension. Such inconsistency suggests that the amount of language contact that students received was probably around a certain threshold that was only enough to influence part of the L2 skills, but not the entirety of them. Students’ L2-related behaviors in the SA context might not always lead to expected outcomes if the quality does not accompany quantity.

2.3.1.2 Measuring language contact during SA

A variety of quantitative and qualitative measures have been employed to measure the intensity of informal L2 contact or to reveal the nature of L2 interaction. In this section, widely-adopted instruments for capturing informal L2 contact will be discussed, from quantitative measures to qualitative measures.

2.3.1.2a The Language Contact Profile (LCP)

Among all the questionnaires measuring informal language contact, the LCP might be

the most widely used (Freed, Dewey, et al., 2004). It was not only adopted by a series of studies in a special issue of *Studies in Second Language Acquisition* in 2004 but also adapted by a multitude of studies in the following decades (Briggs, 2015a; Devlin, 2019; Dewey, Belnap, et al., 2013; Hardison, 2014; Hernández, 2010; Serrano et al., 2016; N. Taguchi, 2008; Trentman, 2013). Some studies altered the name slightly, such as the Language Contact Survey (N. Taguchi, 2008) and the Social Contact Questionnaire (N. Taguchi et al., 2016). The original LCP was developed for L1 English learners of L2 Spanish, with a pre-sojourn and post-sojourn section. The post-sojourn section has 37 questions that require students to estimate the time of engagement in interactive contact (17 questions), reading (9 questions), writing (7 questions), and listening (4 questions). Despite its popularity, the LCP is subject to some criticisms, and a number of changes have been made to the original version over the years.

Figure 2.3 Part of the LCP

3. This semester, outside of class, I tried to speak <i>Spanish</i> to:								
3a. my instructor								
Typically, how many <i>days per week</i> ?	0	1	2	3	4	5	6	7
On those days, typically how many <i>hours per day</i> ?	0-1	1-2	2-3	3-4	4-5	more than 5		

First, even excluding the background questions, the excessive length of the LCP makes it impractical to administer, which has resulted in several researchers shortening it. For example, the number of L2-using items was reduced from 37 to 7 (Devlin, 2019), 10 (Hernández, 2010), 12 (Llanes & Muñoz, 2013), 14 (N. Taguchi et al., 2016), 28 (Briggs, 2015a), and so on, depending on the focus of the studies. Second, how to quantify the L2 use of participants is a core issue. Given the bias of self-report data, the validity of the LCP has been of concern (Fernandez & Gates Tapia, 2016) and the overestimation of the total time is common (Dewey, 2017). Thus, Dewey (2017) summarized that breaking down LCP results and analyzing by items or by categories involving social interaction (i.e., interactive L2 use) was crucial. The original LCP asked students to choose the daily hours by a range and recall how many days in a week they performed those tasks (Figure 2.3). As Briggs (2015b, p. 299) argued, “an out-of-class situation may not present itself regularly”. In other words, the time students spend on the same

activity might differ every day. Scholars have adapted the measurement in different ways. Devlin (2019) used a three-level categorization: ✓ for sometimes, ✓✓ for often, and ✓✓✓ for very often. Hernández (2010) and N. Taguchi et al.(2016) simply asked students to report estimated hours of doing certain activities per week. Briggs (2015b) made a significant modification by turning the scale into a “how true of me” rating to find out which scenarios were most highly identified with by the sample. Serrano et al. (2016) asked about the frequency of L2 use on a five-point scale: *almost never, few times, sometimes, often, almost always*, believing that these scales are less quantifiable but more reliable. It can be seen that scholars were striving to strike a balance between the accuracy of data and the feasibility of administration. More details on how the current study modifies the LCP will be discussed in Section 3.4.3.1.

2.3.1.2b The Language Engagement Questionnaire (LEQ)

The LEQ was designed by researchers for the LANGSNAP project based on the LCP (McManus et al., 2014). The biggest difference of the LEQ from previous instruments is that it views SA as a multilingual experience, so it asks students to nominate their L1, L2, and L3 (if any). Then, respondents need to indicate how often they do the 26 listed activities in each language (*every day, several times a week, a few times a week, rarely, never*) on Qualtrics. It gives students a prompt for frequency recollection which is less affected by individual interpretation. The LEQ is a comprehensive instrument for documenting the language use of SA participants with strong practicality, but not all the information it obtains is relevant to the specific research questions of SA studies. For studies that are only interested in L2 development, the LEQ might be redundant in collecting information on L1 or L3 use. Another limitation is that for English L2 speakers, the semantic difference between “*several times a week*” and “*a few times a week*” might be negligible. Thus, the wording of the scales could be modified to avoid confusion. The LEQ was adopted by some recent studies (Arvidsson, 2019) and Köylü (2016) developed a Language Interaction Questionnaire (LIQ) based on the LCP and the LEQ. Section 3.4.3.1 will mention how the researcher incorporated elements from the LEQ.

2.3.1.2c The language log

Figure 2.4 The language log (Martinsen et al., 2010)

Each day, record the number of minutes that you spend speaking, reading, listening to, or writing in Spanish while engaged in the activities listed below

Language Log

	Tues 5/ 27	Wed 5/28	Thurs 5/29	Fri 5/30	Sat 5/31	Sun 6/1	Mon 6/2
Getting ready for school							
Commuting, Public Transportation/Walking							
Eating breakfast							
In classes							
Eating meals							
Talking to friends/roommate							
Watching TV							
Listening to music							
Preparing meals							
Cleaning							
Studying/Doing Homework							
Email							
Reading							
Talking on the phone							
Sunday School							
Family Home Evening							
Talking to Host Family							
Talking to other native speakers							
Talking to other Americans in Spanish							
Other: specify							
Total	0	0	0	0	0	0	0

To have learners reflect more immediately following their L2 use for an extended period of time, the language log was adopted by some researchers (Brecht et al., 1995; Martinsen et al., 2010). The language log (Figure 2.4) consists of a table in which participants specify the minutes that they used the target L2 on 20 activities in seven days of the week and calculate the total time they used the L2 in a day (Martinsen et al., 2010). It was sent out to the students by email. The advantage of the language log is its clear layout and format, which offers data closer to the actual pattern of language use, because it allows learners to compare the time they spent on certain activities between days and modify them. However, it requires learners to complete it seven times a week,

which increases the time commitment of the participants. Similar to the LCP, retrospective estimation of the exact time spent on these detailed activities might be too demanding for participants, and the accuracy of data might be questionable (Dewey, 2017). Martinsen et al. (2010) and Baker-Smemoe et al. (2014) did not find a significant relationship between the hours of L2 use in the language log and linguistics gains ($p > .422$ and $p > .361$, respectively). The authors of the former study speculated that this result may simply be due to the short-term nature of the program, while the authors from the latter study inferred that the quantity of L2 use does not reflect its quality. Since the researcher could not force the participants in the current study to complete the language log daily, this instrument was not adopted.

2.3.1.2d The Daily Linguistic Questionnaire (DLQ)

The DLQ is another recent attempt to “help participants report time as accurately as possible for each of the four language skills in the L1 and L2 while abroad” (García-Amaya, 2017, p. 64). Created on the Formstack platform, the DLQ could be accessed via computer, tablet, or smartphone, which is convenient for the participants. The grids are arranged in a chronological order in four blocks: sleep and waking up, morning, afternoon, and evening, respectively. In each block, students are prompted by some schedule-related questions and they need to indicate how much time they spend on four types (listening, speaking, reading, and writing) of L1 and L2 use in this block of time (Figure 2.5). The DLQ for class day and no-class day differs slightly, which accounts for the different daily routines of students. The advantage of this instrument is that if completed accurately by participants, it can yield micro-level fine-grained details of their daily L1 and L2 use during the SA period. However, this instrument has two potential limitations. One is that some activities involved more than one L2 skill. For example, using Facebook can involve L2 reading and L2 listening if students read some posts and watch some videos. Thus, whether and how students account for these complications might be inconsistent. Second, there are 11 duration options with different time intervals as short as 5 minutes, up to 30 minutes. Some amalgamations could be considered to simplify the layout of the DLQ. For a similar reason as the

language log, the current study did not employ the DLQ, which relies on participants' self-discipline to complete it daily.

Figure 2.5 Part of Block 1 of the DLQ

	None	1–5 min	5–15 min	15–30 min	30–45 min	45–60 min	1–1.5 h	1.5–2 h	2–2.5 h	2.5–3 h	3+ h
Speaking (Spanish)											
Speaking (English)											
Listening (Spanish)											
Listening (English)											
Reading (Spanish)											
Reading (English)											
Writing (Spanish)											
Writing (English)											

2.3.1.2e *Qualitative measures*

In terms of qualitative measures, observations, journal entries, diaries, and interviews are often employed by researchers (Conroy, 2018; Dewaele & Regan, 2001; Dörnyei et al., 2004; Kinginger, 2008; Kinginger & Blattner, 2009), telling compelling stories of individuals or triangulating quantitative findings to strengthen the robustness of research. Some of these studies have reported the effects of non-interactive L2 contact on L2 development. For example, those who enjoyed watching TV and movies in L2 did not seem to improve their L2 as a result (Conroy, 2018; Magnan & Back, 2007; Rivers, 1998), possibly because the level of engagement was not very high in this activity. Most qualitative studies have focused on elucidating the quality of interactive contact during SA, with two notable themes. One is the large individual variation

among students regarding their attitudes, intercultural awareness, and level of social interaction in the host community. The other general finding is the limited interaction opportunities with host nationals, which was observed with regard to students in independent living arrangements and homestay students. Some studies on the living experience of homestay students showed a different reality which seems to debunk a myth that this cohort receives the most immersive SA experience (Kinging, 2008; Wilkinson, 1998). Due to the ability of qualitative methods to delineate the SA experience, the current study will adopt both quantitative and qualitative measures to investigate the language contact of the participants.

2.3.2 Social networks

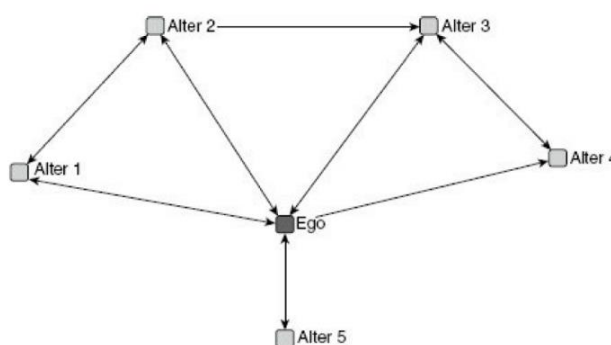
As demonstrated in prior sections, a lack of consistent correlation between L2 contact and L2 gains leads to speculation that the amount of language use might not be the best predictor of L2 gains during SA, and interpersonal interaction might play a decisive role (Baker-Smemoe et al., 2014). Research on social networks can help researchers to delve into the quality of interactive L2 contact of SA participants to explain changes in their language behaviors.

2.3.2.1 Social network research in SA

Social network theory examines the relationships that an individual contracts with others. This theory emerged in sociology and was popularized in the social sciences. Isabelli-García (2006) was one of the first to introduce the construct of social networks to SA research to explain L2 development. In SLA, some sociolinguists used social networks to examine phonological stylistic variations of L2 learners during SA (Gautier & Chevrot, 2015; Kennedy Terry, 2017), while some qualitative researchers examined interaction with host-nationals from the language socialization perspective (Diao, 2016; Wang, 2010). As Dewey et al. (2012) pointed out, both the cognitive and sociocultural approaches lay emphasis on L2 interaction, and there is a consensus that “greater contact with the local community leads to greater gains” (Coleman, 2015, p. 42).

This review will refer to the literature on the social networks of international students in social psychology, international education, intercultural communication, and SLA to explain relevant theoretical notions. The two main kinds of social network analyses are ego network analysis and complete network analysis, with the former focusing on one target individual and the latter on a group of people (Borgatti & Lopez-Kidwell, 2014). SA literature usually conducts ego network analysis. In an ego-net (see Figure 2.6), the subjects are known as “egos”; the people they connect with are called “alters” (i.e., friends, contacts); and the relationships between egos and alters are referred to as “ties” (N. Crossley et al., 2015).

Figure 2.6 Example of ego-net



A vast majority of empirical studies show that SA learners regularly form social groups with their fellow participants or with learners who speak the same L1 as themselves and these “ghettos” are typical of study abroad (Dörnyei et al., 2004; Furnham & Alibhai, 1985; Ife, 2000; Kinginger, 2008; Rienties & Nolan, 2014; Wilkinson, 1998, 2002). Many theories can be drawn on to explain whom international students befriend during SA. Based on the cross-cultural adaptation theory (Kim, 2001), co-national friendship can offer psychological support to SA students in a short run. Besides, *homophily* is a term that means similar individuals tend to move toward each other and act in a similar manner (McPherson et al., 2001). In the SA environment where international students are far away from home, it is understandable that they want to get close to like-minded people or people with similar backgrounds. On the other hand, making local friends requires social capital. By definition, social capital could be

viewed as indirect access to resources or other people (N. Crossley et al., 2015). For SA participants, social capital is “the ability of language learners to secure language and culture benefits by means of access to already established social networks”(Castañeda & Zirger, 2011, p. 548). For instance, McManus et al. (2014) found that students with certain skills (e.g. music, sports, IT skills) were most successful in establishing local friendships, other than those having local romantic partners. These theories could be drawn on to elucidate the social network formation of the current sample.

Another quest for researching the social networks of SA students is the developmental trajectory. However, most of the detailed changes were recorded through qualitative case studies (Isabelli-García, 2006; C. Li et al., 2021), as quantitative studies tended to reveal group-level trends in the friendship pattern of SA participants (Dewey, Belnap, et al., 2013; Dewey et al., 2012; Gautier & Chevrot, 2015). Coleman concluded that the social networks of SA learners “tend to be formed early, and may subsequently fossilize or develop” (2013, p. 31). For example, McManus et al. (2014) found that British learners in France had smaller L1 and L2 networks at the end of a one-year residence abroad, suggesting a shrinkage in the size of overall social networks. Among the five closest friends the participants listed, the proportion of L2 speakers did not increase over the year. Some masters theses identified that foundation students’ social networks in the USA tend to decrease in size but increased in quality (Dewey, personal communication, July 19, 2022), and similar research could be done in the British context to unravel the longitudinal changes in the L2 social networks.

Using social networks to explain L2 development has been gaining traction in recent years. So far, quantitative and qualitative research in this line points to a generally positive relationship in oral proficiency (Baker-Smemoe et al., 2014; Dewey, Belnap, et al., 2013; Xie & Liu, 2021), oral fluency (Bejarano et al., 2019), sociolinguistic variations (Gautier & Chevrot, 2015; Kennedy Terry, 2017), pragmatic choices (C. Li et al., 2021), lexical complexity (McManus, 2019) and self-perceived writing proficiency (Strobl & Baten, 2021). While the relationship between social networks and

oral proficiency gains could be observed within a semester (Baker-Smemoe et al., 2014; Gautier & Chevrot, 2015), social networks and writing development did not show a significant relationship over one term (Strobl & Baten, 2021). Similarly, lexical complexity did not show a significant relationship with social network variables until the end of one academic year (McManus, 2019).

Further, not all aspects of social networks were significant predictors. To be specific, the *intensity* and *dispersion* of social networks strongly predicted the perceived L2 gains of American students studying abroad in the Middle East, Asia, Europe, and Central America (Baker-Smemoe et al., 2014; Dewey, Belnap, et al., 2013; Dewey et al., 2012). On one hand, this result underscored the importance of intense relationships, which are likely to involve in-depth extended conversations on varied topics. These conversations were more beneficial to L2 acquisition than rote and superficial conversations (Yager, 1998). On the other hand, network dispersion matters because the diversity of friends' backgrounds provides learners with linguistically rich L2 exposure, which enables them to leverage non-redundant information about normative practices in different communities (C. Li et al., 2021). In contrast, the *size* of social networks did not play a central role in L2 development during SA. If anything, a decreased network size predicted greater proficiency gains (Baker-Smemoe et al., 2014). They speculated that this change implied that participants devoted more time to emotionally close friends and spent less time with weak ties (Granovetter, 1973). For the current study, it is necessary to consider the different features of social networks to fully understand their relationships with PV acquisition.

2.3.2.2 Measuring social networks during SA

Quantitative and qualitative measures have different strengths in capturing the social networks of SA participants. In quantitative research, surveys are a mainstream method for collecting social network data, as data are collected under relatively standardized conditions and only require modest time commitments from participants (Scott & Carrington, 2011). Social network surveys are usually developed in different forms of

‘name generators’ to elicit information about the ego-networks. The present study designed a new instrument based on a few existing tools and each of them is discussed in turn.

2.3.2.2a The Study Abroad Social Interaction Questionnaire (SASIQ)

The SASIQ (Dewey, Belnap, et al., 2013; Dewey et al., 2012) is an influential 13-item questionnaire that uses a name generator to evaluate the characteristics of the learners’ social networks, which was developed based on previous instruments. The questionnaire asks the respondents to name up to 20 contacts whom they speak L2 with and choose from a few drop-down menus to describe the characteristics of their contacts. Some questions documented the proportion of L1 and L2 use with their friends. It also has some open-ended questions, such as how they met, the reasons why they were good friends or just acquaintances, and how the students make local friends. The key data range from the size (the number of contacts), intensity (their closeness), durability (the frequency of interactions), density (the size of social groups), to dispersion (the number of social groups a respondent has) of the network, and the English proficiency of their friends.

Although the five dimensions above can capture the L2 friendship patterns of learners to a large extent, the SASIQ cannot show “the nature of interactions they actually engage in” (Dewey, Belnap, et al., 2013, p. 102). In other words, it cannot reveal exactly what they do with their L2 friends. In addition, the notion of *durability* could potentially be replaced by the *frequency* of contact, as in essence “how often a person associates with an individual” (Dewey et al., 2012, p. 120) denotes their frequency of contact. Finally, the SASIQ only collects information regarding face-to-face interaction with L2 speakers. Online interaction was not included, which limits our understanding of the overall social interaction of the sojourners.

2.3.2.2b The Social Networks Questionnaire (SNQ)

The LANGSNAP group utilized an SNQ to record interpersonal relationships in five different social contexts (McManus et al., 2014). For each context (work/university,

organized free time, general free time, home, virtual social activities), participants list people with whom they interact regularly over the past month and specify the frequency of contact, language use, the relationship, and the occasion of the first contact (Figure 2.7). Finally, students are required to nominate the five closest contacts across all contexts. The SNQ can reveal a comprehensive picture of the multilingual social relations of SA participants because it incorporates the virtual environment and it is not limited to L2-speaking friends only. One limitation of the SNQ is that it does not collect information about the relationship between the alters, so researchers do not know the density and dispersion of the networks. Moreover, similar to the SASIQ, it does not obtain information on what the respondents do together with their friends. Furthermore, if the respondent is close to someone in more than one context, that person's information has to be put in several times, which might cause fatigue for respondents.

Figure 2.7 Screenshot of the SNQ

4. Work/Uni Context, Person 1

These questions refer to the person below:

Kevin

Question 1.

How often do you interact with this person?

☐ every day

☐ several times a week

☐ a few times a week

☒ a couple times a month

Question 2.

What language(s) do you use when communicating with this person?

Please select ▼

Question 3.

What's your relationship to this person?

Question 4.

How did you first meet? (e.g., through a mutual friend, at work, etc. -- just a brief response)

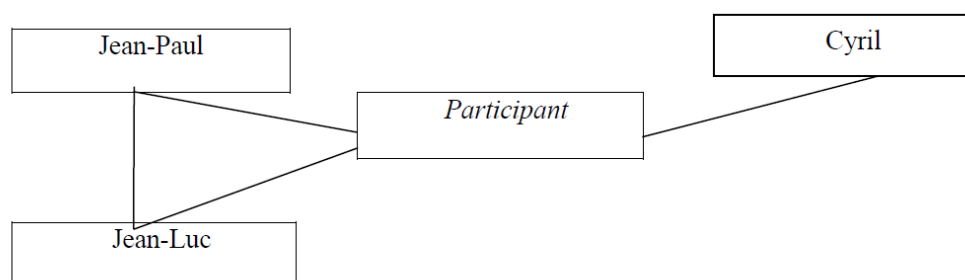
2.3.2.2c *The Social Network Strength Scale (SNSS)*

Kennedy Terry (2017) adapted the original SNSS (Milroy & Milroy, 1978) for the SA context, focusing on two density and two multiplexity measures of L2 social networks. Density measures attempt to gauge the intensity of interaction with L2 native speakers, while multiplexity measures are designed to capture the richness of the learner's L2 interactions by investigating what the learner does and what the learner talks about with native speakers. To be specific, Density Measure 1 multiplies the number of friends listed by the respondents and the number of hours that they speak. Density Measure 2 calculates the number of ties between the alters (Figure 2.8). Multiplexity Measure 1 focuses on the range and frequency of social activities in L2, while Multiplexity Measure 2 calculates the breadth of topics the respondent discussed in L2. In this way, this instrument taps into the quality of social interaction using multiplexity measures. However, this instrument does not probe into the profile of the learners' alters (except for age and their relationship). Second, interactions with non-native L2 speakers are not accounted for. Further, the hand-drawn density measure is impractical for the administration and analysis of a large sample.

Figure 2.8 Part of the SNSS

Question #2 (Density Measure #2)

Now use the names listed above to draw a social network grid. Place your name in the first box, the name of each person in a separate box, and connect the boxes with a line to show who knows whom. An example is provided for you:

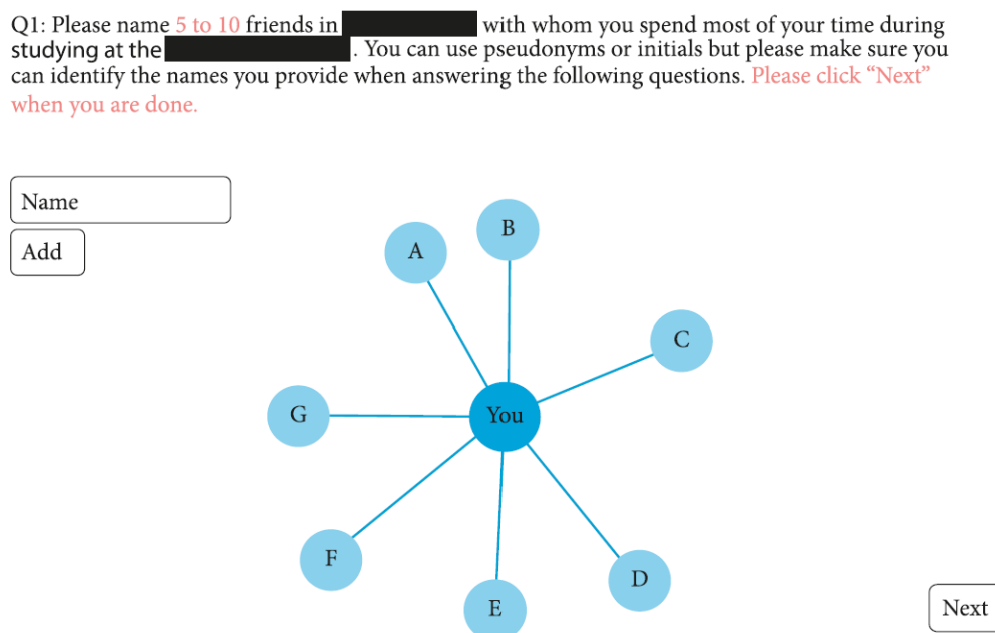


2.3.2.2d *The personal network questionnaire*

The personal network questionnaire (Xie, 2017; Xie & Liu, 2021) was adapted from GENSI (Graphical Ego-Centered Network Survey Interface) (Stark & Krosnick, 2017).

This questionnaire has an interactive web-based interface in which the nominated friends appear with ties to the ego (Figure 2.9). This questionnaire collects five social network variables: the size of networks (the number of friends), the intensity of relationship (the strength of relationship), frequency of contact, the density of networks (the proportion of existing ties in total ties possible), and the similarity of networks (whether there were L1-speaking or English-speaking contacts). A unique feature is the similarity parameter, which examines whether the networks of participants are mono-cultural or cross-cultural. However, this questionnaire only accounts for face-to-face contact and did not probe into the specific nationalities of the contacts. Besides, the codes of this tool were not publicly available, thereby making reuse and adaptation of this tool difficult.

Figure 2.9 Screenshot of the personal network questionnaire



Based on the above review, the researcher developed a Study Abroad Social Network Survey (SASNS) for the current study. The SASNS only concerns data of L2 social networks and omits questions on L1 social networks, given the limited patience of the participants and the relatedness to the research question. Further, the SASNS aims to

collect comprehensive information on the L2 network, including detailed friends' profiles and the diversity of topics and activities the respondents engage in. Table 2.1 displays the pros and cons of some existing social network questionnaires, and Section 3.4.3.3 will delineate how the SASNS merges the merits of existing tools.

Table 2.1 Comparison of existing social network questionnaires

Instrument	Variables	Strengths	Limitations
The SASIQ	◆ Size ◆ Intensity ◆ Durability ◆ Density ◆ Dispersion	Comprehensive evaluation of the quality of L2 social networks	No data on activity multiplexity and topic multiplexity
The SNQ	◆ The number and characteristics of friends	Comprehensive view of social networks in five contexts in L1 and L2	No data on network density or dispersion
The SNSS	◆ Density ◆ Multiplexity	Includes activity multiplexity and topic multiplexity	Only includes interaction with L2 native speakers Limited information on the alters
The personal network questionnaire	◆ Size ◆ Intensity ◆ Frequency ◆ Density ◆ Similarity	The interactive web-based interface for network density	No data on activity multiplexity and topic multiplexity

2.3.2.2e Qualitative measures

Qualitative measures (e.g. interviews, journals and diaries, observations) play a vital role in researching the social networks of SA participants because qualitative data can not only triangulate quantitative data to shed light on the large variations in the SA experiences of a sample (Dewey, Ring, et al., 2013; Trentman, 2013; Wilkinson, 2002) but also explore the fine-grained nuances of interpersonal relationships that affect the L2 development of individuals (Dörnyei et al., 2004; Isabelli-García, 2006; Kinginger & Blattner, 2009). From these studies, it is unsurprising that participants with improved

post-test performance spoke the L2 more often with native speakers or other international friends, while participants who did not improve their L2 mostly socialize with their co-nationals, which suggests the crucial role of L2 interaction and the interplay between learner attributes and external environments.

Methodology-wise, either L1 or L2 have been used in semi-structured interviews to explore social integration. The studies that used L2 for interviews usually had a dual purpose of collecting L2 speech samples for gain analysis and obtaining commentaries on their SA life (Adolphs & Durow, 2004; Magnan & Back, 2007; Tracy-Ventura, 2017). However, this method seems more suitable for advanced L2 speakers, such as the language-major participants in the LANGSNAP project, because the participants' self-presentation in the interviews was affected partly by their personal confidence as L2 users (Mitchell, 2015). Other cases of using L2 for interviews usually happen when the interviewer and the interviewees did not share the same L1 (L. Brown & Holloway, 2008; Dörnyei et al., 2004; Wright & Schartner, 2013). When the interviewer and interviewees had the same L1, using L1 was preferred for open and relaxed self-expression and avoidance of language barriers, especially for Asian students (Spencer-Oatey & Xiong, 2006; Tanaka, 2007; Trentman, 2013; Xie, 2017). In the current study, interviewees could choose the interview language where linguistically possible (Chinese/English).

2.3.2.3 Social network models for international students

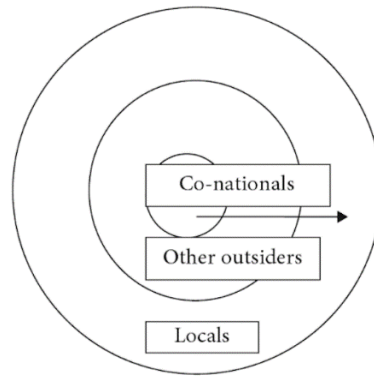
Because the SASNS collects the nationality data of the L2 contacts of the sample, it is possible to identify where their L2 friends mostly come from. In social psychology, Bochner et al. (1977) developed a functional model of the social network of international students. In this model, the friends of international students were categorized into three subsets: co-nationals, host-nationals, and multi-nationals. The primary, mono-cultural network is made up of close friendships with compatriots. Co-national friends can help sojourners ease nervousness and homesickness during SA and affirm their own cultural identities. The secondary bi-cultural network consists of host-

nationals who can offer instrumental facilitation of academic and professional aspirations, such as obtaining information and improving their language. Finally, a third and multicultural network comprised of friends from other countries. The multi-national friends share the same challenge and foreignness as the international students, and they can provide companionship in recreational activities in SA.

However, Bochner et al. (1977) only recruited 30 participants in Hawai'i, so their findings that participants had more local friends than international friends might no longer be contextually representative of the internationalized higher education today. In a replication and extension study on 140 participants in the UK (Furnham & Alibhai, 1985), the listed international friends outnumbered the local friends, which countered the finding in Bochner et al. (1977). Still, in terms of the order of desirable friendship, Furnham and Alibhai (1985) provided empirical support to the functional model (Bochner et al., 1977) that friendship with host-nationals was more desirable than that with multi-nationals.

Coleman (2013) put forward a concentric circles model describing the dynamic socialization patterns of SA sojourners (see Figure 2.10). The three circles are labeled co-nationals, other outsiders (i.e. multi-nationals), and locals (i.e. host-nationals) from the inner to outer layers. Coleman (2013) holds that the progression toward the local community is not universal or automatic, but given motivation, time, and effort, movement tends to be centrifugal. It can be seen that the order of multi-nationals and host-nationals reverses that in the functional model by Bochner et al. (1977) but it is in agreement with extensive SA studies (Furnham & Alibhai, 1985; Schartner, 2015).

Figure 2.10 The concentric framework of Coleman (2013)



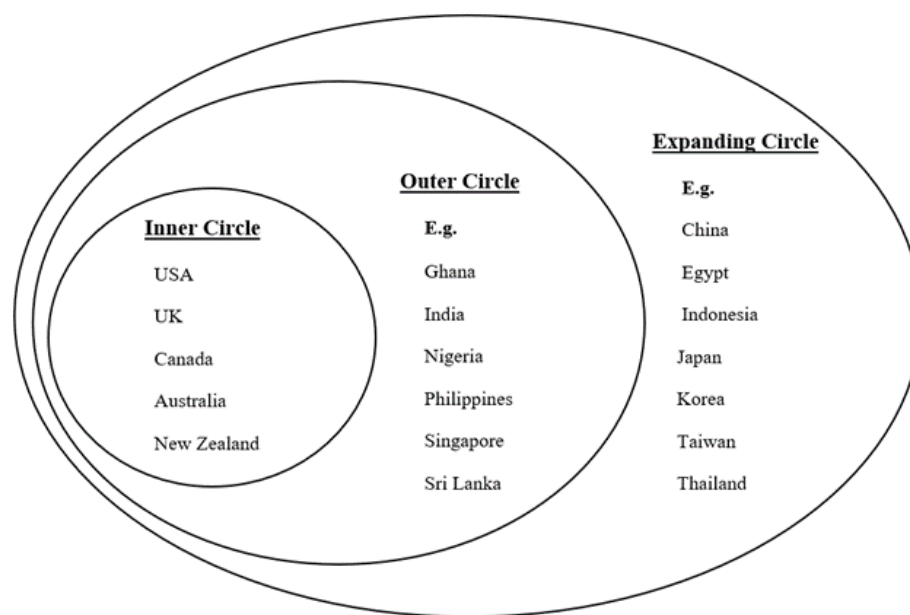
Coleman (2013, p. 31) recognized that his model “ignores individual and contextual variation, and over-simplifies reality”. He had extensive research experience with British students studying other European languages abroad. The social status and mindset of a Chinese student in the UK are probably inherently different from those of a British student in France, as the latter has a more equal social capital and closer cultural distance. Thus, for international students in different host countries, tailored modification is needed case by case. In the current study, the power relationship and cultural distance between countries should be considered.

Regarding power relationships, the center-periphery (or core-periphery) model (Friedmann, 1966) is implicated in various world system theories (Wallerstein, 2004). In this system, the US and most European countries are considered core nations, whereas most countries in Latin America, Asia, and Africa are considered peripheral (Diniz & Noels, 2021). In a global context, core nations hold power in the economy, human capital, and knowledge construction, attracting many international students from peripheral countries (Diniz & Noels, 2021). Thus, students from peripheral countries might have a disadvantage in their social capital compared with students from other core countries when studying in the UK.

As for cultural distance, language is an important parameter of the mental distance of international students in the UK. English as a lingua franca, is the L1 of many countries because of its colonial and emigration history. Kachru’s Three Circles of English model (1982) (Figure 2.11) demonstrated that, commonly, people from the Outer Circle countries speak fluent and near-native English because of British colonial history, while

people from the Inner Circle are native English speakers because of ancestral emigration from the UK. Thus, these English native-speaking international students might have an emotional or ancestral tie to the UK and a competitive edge in social integration compared with people from the Expanding Circle, who are L2 speakers of English.

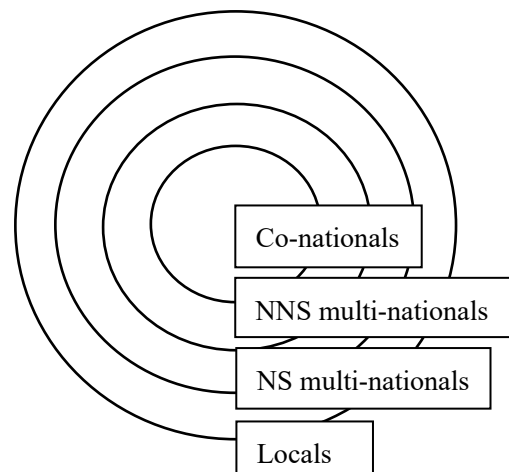
Figure 2.11 Kachrus' Three Circles of English (1982)



Based on the functional model by Bochner et al. (1977), the concentric framework of Coleman (2013), and the three-circle model of Kachru (1982), the present study puts forward an SA social network model for the Expanding Circle students in the UK (Figure 2.12). The four circles are labeled co-nationals, non-native English-speaking multi-nationals, native English-speaking multi-nationals, and host-national networks. The innovation of this adaptation is that within the category of multi-national friends, it is hypothesized that students might form closer relationships with non-native L2 speakers just like themselves. The literature suggests that friends from different circles provide varied support for the adaptation of international students abroad (Kim, 2001). However, the ease of establishing native English-speaking and non-native English-speaking friendships might be different. The literature often reports that limited L2 proficiency is a crucial reason that impedes international students' interaction with L2

native speakers and language anxiety persists when they speak with L2 native speakers (L. Brown, 2008; Gao, 2017). Whether English proficiency matters in friendship formation with international peers can be addressed by the current study.

Figure 2.12 SA social network model in the UK



Note: NNS=Non-native English speaking. NS=native English-speaking

2.4 Investigating phrasal verbs and informal language contact

Extant empirical studies tend to yield mixed findings on the relationship between language contact and L2 gains, whereas studies involving social networks often reveal significant predictors of the L2 gains. Admittedly, these studies differed in terms of L2 competencies being measured, the sample characteristics, and the features of the SA program. Overall, it can be inferred that the quality of L2 use matters more than the quantity of L2 use in L2 development during SA. Although there is a dearth of quantitative studies on the relationship between informal language contact (including social networks) and PVs, a handful of studies have examined the relationship between social interactions and other types of formulaic language.

2.4.1 Formulaic language development and language contact

Case studies revealed that having social interaction with L2 native speakers offers gains in L2 colloquial words (Dewaele & Regan, 2001), formulaic sequences (Adolphs &

Durow, 2004), and colloquial phrases (Kinging & Blattner, 2009), with sample sizes from two to six participants. In quantitative studies, findings to date are highly mixed as to whether language contact is positively related to formulaic language in SA contexts (Arvidsson, 2019; Bardovi-Harlig & Bastos, 2011; N. Taguchi et al., 2013).

For example, Arvidsson (2019) recruited 41 L1 Swedish students studying in France for 12-18 weeks to complete an LEQ and a self-developed French multi-word expression test. Neither the total amount of out-of-class language contact nor the amount of interactive out-of-class contact significantly predicted multi-word expression gains. In the same vein, N. Taguchi et al. (2013) tracked the Chinese formulaic expressions development of 31 American students who studied in China for one term. They adopted a Formulae Contact Questionnaire and a computerized Discourse Completion Task. Correlation analysis revealed no significant relationship between formulaic expression gains and the reported frequency of relevant scenario encounters. In regression analysis, the frequency of encounters had a small but significant effect only for the lower-proficiency learners. These two studies had rigorous research designs, such as having multiple raters for the tests and taking initial L2 proficiency as a confounder. However, they did not report qualitative data on language contact, so the researchers' interpretations of the non-significant relationship were largely based on existing literature and their inferences, rather than the accounts of the participants.

In another study, Bardovi-Harlig and Bastos (2011) recruited 122 ESL learners in the US to complete an aural recognition task and an oral production task of conventional expressions (e.g. *No problem. Watch out!*). They administered a self-developed language contact questionnaire that only concerns speaking and aural input, but not reading and writing contact because they believed the target expressions usually appear in conversations. It should be recognized that their questionnaire had an inconsistent design: when quantifying the amount of English speaking with native speakers, weekly hours were asked; in quantifying English speaking with international friends, daily hours were asked; and to quantify American TV input, only the frequency was asked

(e.g., once a week, once a day). Fortunately, in the analysis, the researchers converted this data to scales and added up the sum of engagement levels in each activity, which minimized the influence of different quantifications. Regression analysis demonstrated that intensity of interaction had a significant effect on the recognition ($p < .01$) and production of conventional expressions ($p < .05$).

These contradictory findings may be attributed to two reasons. Regarding research design, the former two studies took a pre-test and post-test longitudinal approach, while Bardovi-Harlig and Bastos (2011) adopted a one-off cross-sectional design. The longitudinal tracking periods were possibly not sufficient in length for any relationship with gains to be detected (Briggs, 2015a) (i.e. because the amount of contact and/or the magnitude of gains made was negligible). Participant-wise, Bardovi-Harlig and Bastos (2011) had the largest sample size and their participants came from four different level classes in an Intensive English Program with many L1s. The participants in Arvidsson (2019) and N. Taguchi et al. (2013) had homogeneous L1 and cultural backgrounds, which might have led to less variance in the language contact reported and thus to the lack of significant results. The current research design follows that of Bardovi-Harlig and Bastos (2011), involving a larger sample of learners of varied proficiency and L1s, and adopts a longitudinal approach such as Arvidsson (2019) and N. Taguchi et al. (2013), to delve into the relationship between PV knowledge, PV gains, and L2 contact.

A few studies examined the formulaic language development of SA participants but did not statistically examine the impact of language contact directly (Alqarni, 2017; Schmitt, Dörnyei, et al., 2004; Serrano et al., 2016). Alqarni (2017) administered a 12-item collocation test to 124 Saudi students in Australia. This study found that students with longer lengths of stay in Australia had higher collocation scores. Descriptive statistics showed that the students with one to three years of stay had approximately the same scores, while the four-year-stay group had much higher mean scores. The participants were enrolled in different Australian schools in different states across the country, but no discussion was provided by the author regarding whether their living arrangements and school curricula differed. Serrano et al. (2016) carried out a formulaic

sequence test on an SA group and an at-home intensive group of Spanish teenagers learning English. This 16-item judgment test was designed with the hypothesis that the SA learners would outperform their peers at home because they can practice a lot of formulaic expressions (e.g., *How are you today? How much is this?*) with L2 speakers. The results showed that both groups made significant gains in three to four weeks, and the SA group outperformed the at-home group. This study did not report the statistical relationship between language contact and formulaic sequence test scores, but the authors concluded from the language contact questionnaire results that the closest friends of most SA participants were L1-speaking, and not all learners took advantage of all the opportunities they had to use the L2 during SA.

In a large-scale longitudinal project led by Schmitt (2004), 94 international students at an English university completed a cloze-format productive formulaic sequence test and a multiple-choice receptive formulaic sequence test before and after a three-month pre-session course. A VLT 3000 and a VLT 5000 (i.e., single-word vocabulary knowledge measures) functioned as baseline measures. Two questionnaires on language aptitude and attitude/motivation and a series of semi-structured interviews were also completed. It was found that both receptive and productive formulaic sequence tests saw significant development, although the receptive test had a *ceiling effect* with nearly half of the participants (34/70) obtaining a full score in the post-test (Schmitt, Dörnyei, et al., 2004). Surprisingly, learner attributes, such as language aptitude, attitudes, and motivation, did not show a statistically significant correlation with formulaic sequence gains. A case study of the subsample showed that learners who had intensive interaction with British native speakers or international friends achieved the greatest improvement in formulaic language, and their low-scoring counterparts mostly interacted with their co-nationals (Dörnyei et al., 2004). Their findings highlighted the pivotal role of sociocultural adaptation (acculturation) in the formulaic language development of L2 learners. As the sample of this project resembles the participants in the current study, many elements in the procedure design of Schmitt, Dörnyei, et al. (2004) will be adopted.

Lastly, Conroy (2018) is an exception that explored the relationship between living experiences in SA (contextual factors) and the orthographic command of idiomatic PVs. 41 Chinese L1 university students completed a sentence dictation task before and after a six-week homestay SA program in Australia to test their form recall of PVs. Students were given one point each for the correct verb and the correct particle. The chosen PVs were highly colloquial and were not taught in class (e.g. *hop on*, *pop in*). The results illustrated that most learners made very significant improvement on idiomatic PVs after this program ($t[40] = 9.08, p < .001$). Qualitative data were collected from journals and diaries, implying that quality social interaction (“lots of opportunities to interact with a range of native speakers”) facilitated PV gains (Conroy, 2018, p. 323). In particular, some learners appreciated the explicit corrective feedback and modified input from their interlocutors. It should be noted that using the same target sentences in the post-test only five weeks after the pre-test, this study might still be subject to some *practice effect*. In addition, this study did not take learner attributes into account (e.g., L2 proficiency, motivation) in explaining their gains. Lastly, the author did not collect quantitative data on informal language contact, which to some extent impaired the argument between the link of homestay experiences and PV gains because L2 contact could not statistically be explored.

2.4.2 Rationale for the current study and original contributions

Based on the review above, the selection of PVs as the dependent variable can be attributed to their importance for communication with English L1 speakers and their difficulty for L2 learners. The rationale for conducting the current study on SA participants is driven by the tenets of usage-based approach that more L2 contact is likely to lead to more gains. It should be noted that “L2 contact” in the current study embodies various forms and entails both interactive L2 use and receptive L2 exposure, though interactive L2 use may play a more directly beneficial role in L2 gains than receptive L2 exposure. A plethora of studies have confirmed that vocabulary can be acquired incidentally through L2 exposure. For example, incidental learning took place

on one-word vocabulary and formulaic sequences through L2 reading (S. Liu, 2018; Pellicer-Sánchez, 2017; Teng, 2016), watching L2 videos (Majuddin et al., 2021; Vu & Peters, 2021), and watching L2 TV (Puimège & Peters, 2019), mediated by item-related and learner-related factors. Although most of these studies were conducted in an EFL experimental setting, there was also evidence of significant relationships between formulaic language (including PVs) and L2 leisure activities, such as reading in English and social networking in English (Garnier & Schmitt, 2016). Therefore, examining the relationship between out-of-class L2 exposure and PV gains might reveal new insights into language development in an SA context for the first time.

A few original contributions can be made by the present study. First, as Coleman (2013, p. 17) commented, SA research to date “disproportionately addresses certain departure zones, notably North America and Europe”. Less is known about the L2 gains of degree-seeking students from developing countries in highly developed Anglophone countries. The current study can address the scarcity of queries in this line and enrich our understanding of the L2 development of mixed-L1, especially Chinese students in Anglophone contexts like the UK, which the introduction chapter has indicated to be the largest population of international students. In terms of sample size, due to the nature of SA programs, past research typically had a sample size of fewer than 50 students. The current study recruited over 100 participants, which can strengthen the representativeness of research findings on foundation students and will allow for more nuanced data analysis.

Regarding the tracking period, past research has often been undertaken on semester-long exchange students. The foundation program enables the researcher to follow the students for one academic year, which is desirable in SA research due to the incremental and non-linear path of vocabulary development (Fitzpatrick, 2012). Previous studies have shown mixed results on how long it takes for vocabulary gains to be captured in SA students. For example, Serrano et al. (2012) witnessed a significant gain in oral lexical richness after one term of SA, while Mora and Valls-Ferrer (2012) found no significant change in spoken lexical diversity after more than one year of formal

instruction and SA. Contradictions as such highlight the need for additional research on PV development.

Additionally, using multiple instruments in three rounds of data collection will push new boundaries in PV research in two ways. On one hand, this study pioneers in tracking learners' PV development in a naturalistic environment for an extended period. Existing PV studies usually adopt cross-sectional designs or longitudinal experimental designs with interventions. The only study that examined PV development in the SA setting had a short tracking period of six weeks and only assessed orthographic mapping (letter-sound connections) of 10 to 16 idiomatic PVs (Conroy, 2018). On the other hand, this study will shed light on a comprehensive picture of learners' PV development in SA by combining three measures of PV assessment: receptive tests, productive tests, and spoken learner corpora. Therefore, more insight could be obtained regarding the level of mastery of different PVs. Kamarudin et al. (2019) analyzed PV use in a large national learner corpus, which is equivalent to but not the same sample that completed the PV test in their study. Thus, such a comprehensive assessment is only possible through collecting speech samples and test data from the same learner cohort.

Lastly, another original contribution of the present study is using social network analysis to explain PV development, which responds to the call to explore the role of social networks in vocabulary development during SA (Zaytseva et al., 2018). SA benefits vocabulary development because it has the potential to provide input and output opportunities for incidental learning. Social networks, as an indicator of the quality of L2 interaction, have relationships with several L2 competencies (Baten, 2020; McManus, 2019) and this study can fill a gap in research on its impact on PV development. Furthermore, this study developed a new web-based tool to capture the L2 social networks of the participants, which is a methodological innovation.

CHAPTER III METHODOLOGY

3.1 Research questions

Based on the literature review, the present study aims to answer the following research questions (RQs):

1. What is the developmental profile of phrasal verb (PV) knowledge among international foundation program students in the UK over one academic year?
 - 1a. What non-exposure factors influence the PV scores of the participants?
2. What are the patterns and the developmental profile of informal L2 contact among international foundation program students in the UK over the first half of the academic year?
 - 2a. What personal background factors influence the language contact of the participants?
 - 2b. How did the COVID-19 lockdown in the UK in Spring 2020 affect participants' informal language contact and SA experience?
3. What linguistic factors (corpus frequency, semantic opacity), personal factors (overall English proficiency, L1), and L2 exposure (language contact, social networks) influence PV development in these participants during study abroad?

As demonstrated in Chapter Two, there is a broad spectrum of “knowing” a word (Schmitt, 2014). In this study, PV knowledge is operationalized as the knowledge of form-meaning mapping of PVs in the tests and the appropriate use of PVs in the oral elicitation task. Informal language contact is operationalized as the amount, type, and quality of the time students are exposed to/using English outside the classroom. As shown in Section 2.3.2, the reason social networking data are included in measures of language contact for the current study is that it captures the quality of spoken L2 contact students had, which adds rich details of the language contact of the participants.

3.2 Study design

The present study employs an explanatory simultaneous QUAN + qual design (Morse, 2010). The uppercase and lowercase distinction suggests that the quantitative component is supplemented by the qualitative component. Mixed methods have gained momentum in SLA in recent decades, enabling researchers to report holistic pictures of learner cohorts as well as fine-grained details of individual differences (Watson & Wolfel, 2015). Very often, a single method is insufficient to answer complex research questions, whereas mixed-methods studies can deploy multiple methods to address that limitation. The quantitative measures in the present study encompass a variety of questionnaires and tests on informal language contact and PV knowledge of foundation program students in the UK, as well as their general vocabulary knowledge at the beginning of the sojourn (Table 3.1). Meanwhile, the qualitative measures include administration of the IELTS speaking test (to yield a spoken learner corpus) and semi-structured interviews with a subsample of the participants.

Table 3.1 Simultaneous mixed-methods study design

Time points	Time 1 (T1) 8 weeks after arrival 11-19 Nov 2019	Time 2 (T2) +12 weeks 17-26 Feb 2020	Time 3 (T3) +10 weeks 1-15 May 2020
Quantitative measures	Receptive and productive PV test	Receptive and productive PV test	Receptive and productive PV test
	Vocabulary Levels Test (VLT) 5k	VLT 5k	/
	Language contact questionnaire (LCQ)	LCQ	LCQ
	Study-abroad social network survey (SASNS)	SASNS	SASNS
Qualitative measures	IELTS speaking test	IELTS speaking test	IELTS speaking test
	Semi-structured interview	Semi-structured interview	Semi-structured interview

The strengths and limitations of the main instruments are discussed below. The popularity of questionnaires in L2 research can be attributed to their easiness to

construct, versatility, and unique capability of gathering a multitude of information quickly, which saves the researcher time, researcher effort, and financial resources (Dörnyei & Taguchi, 2009). The disadvantages of questionnaires, however, include unreliable and unmotivated respondents, *prestige bias*, no opportunity to correct respondents' mistakes, simplicity of answers, and so on (Dörnyei & Taguchi, 2009). Interviews are useful in investigating individuals' experiences, beliefs, and self-reports of behaviors. An advantage of interviews is that they can explore issues in-depth to cast further explanatory insight into survey data. Nevertheless, limitations of interviews include substantial time investments, *interviewer bias*, inconveniencing respondents, and interviewee fatigue (Cohen et al., 2017). The current study employs standardized speaking tests such as IELTS. The advantage of using rigorous proficiency assessments to elicit speech samples is that they are carefully scrutinized for their reliability and validity, whereas a limitation of large-scale tests like IELTS might lie in their insensitivity to small, incremental changes in performance (Faitaki & Murphy, 2019). The instruments and the corresponding variables they yielded are listed in Table 3.2.

Table 3.2 Research questions and corresponding data

Research questions	Instruments	Data yielded
RQ1 PV knowledge	PV tests, IELTS speaking test	PV test scores, PVs produced in the learner corpus
RQ2 informal language contact	LCQ, SASNS, interviews	Language contact types and amount; social networks quality
RQ3 relationship between PV and language contact	PV tests, IELTS speaking test, LCQ, SASNS, interviews	PV test scores, learner corpus, language contact types and amount; social networks quality

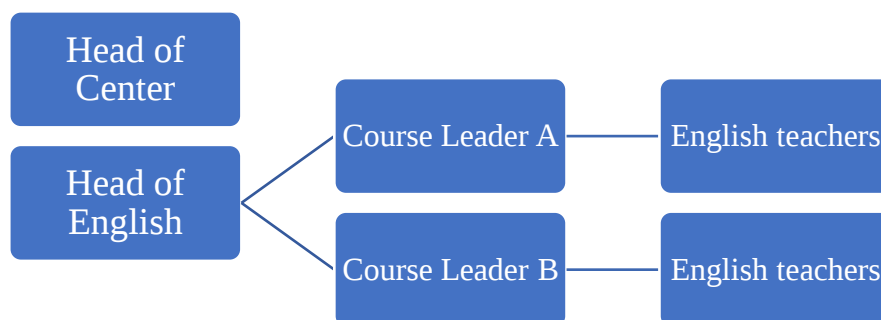
3.3 Sample

3.3.1 Sampling strategy

Overall, the current study opted for a *non-probability sampling* of the student

population (Cohen et al., 2017). The sample was accessed via contact with the Head of English of an education group that runs 15 foundation program centers (independent teaching units on university campuses) across the UK and places over 3000 international students in UK foundation programs annually. Two centers with the largest student numbers in the group (600-700 students) were chosen as suitable research sites by the Head of English for the researcher to recruit about 125 students from each center. Since each class has 15-20 students, the course leaders of English (a role that supervises teachers) and the researcher chose seven three-term classes of B1-B2 level students from each center based on class availability (Figure 3.1). The sample consisted of many Chinese students, with participants from a mixture of other countries too (see Section 3.3.4 for sample breakdown).

Figure 3.1 Liaison hierarchy of the data collection



Since it is a mixed-method study, *sequential mixed-method sampling* was adopted, in which the qualitative sample was selected from the quantitative sample (Cohen et al., 2017). In quantitative data collection, students were encouraged to sign up for the interviews and the researcher introduced some non-financial perks of becoming an interviewee (e.g., a free Oxford tour, English learning advice). Cash bonuses were not provided like in some other SA studies (Arvidsson, 2019; Mitchell et al., 2017) due to limited financial resources. After collecting students' contact details from the questionnaire, the researcher added them on WeChat (for Chinese students) or WhatsApp (for non-Chinese students). After the initial greeting, the researcher asked about each student's availability during the week of her visit and scheduled a time for

an interview.

3.3.2 The foundation programs

University foundation programs are typically one-year courses that equip international students with necessary English/academic preparation for future bachelors or masters studies in the UK. In these foundation programs, 20-25 hours of classes (e.g., English, subject research project, postgraduate skills & development, and tutorials) were scheduled every week during term time. Students attended academic English classes based on their levels and attended subject classes based on their majors (e.g., humanities, engineering, finance & accounting). Cambridge English Empower Student's Book B2 (Doff et al., 2015) was used in English classes. After scrutinizing this book, some useful PVs appear in vocabulary exercises, but the PVs are not accompanied by definitions.

The two research sites are referred to as Center A and Center B and the participant number breakdown is listed in Table 3.3. University A is located in central London. Center A claimed that its foundation curricula were recognized by some top universities in London and it has six partner universities in London. Most students studying at Center A applied for degrees in the universities of a higher ranking than University A. University B is located in a small city in the Midlands of England. The university has a short history and has a large proportion of international students. All the students at Center B can continue their studies at University B provided that they pass all the in-house exams.

Table 3.3: Participant number across sites

Program	Pre-bachelor	Percentage	Pre-master	Percentage	Total
Center A	88	71.5%	35	28.5%	123
Center B	79	60.8%	51	39.3%	130
Total	167	65.4%	86	34.6%	253

3.3.3 Sample characteristics

Regarding sample characteristics, counting from valid percentage, 65.5% of the participants came from China, taking up the lion's share of the sample. The Chinese students in the sample came from different cities in China and applied to the target universities through different student agencies. Other students mostly came from Asia (e.g., Southeast Asia, Central Asia, South Asia, Middle East) (25.5%), with a few from Africa (6%) and Eastern Europe (3%) (see Table 3.4). Although the sample is over-representative of Chinese students, the Chinese student population is of particular research interest for two reasons. About one out of three non-European international students in the UK is Chinese (Study International, 2018), but are under-represented in current SA literature, so their SA experience is worth close examination. Asian students are observed to be shy and reserved in cross-cultural communication and the cultural distance between China and the UK is far (Gao, 2017), so it is important to investigate to what extent students from such backgrounds make use of L2 contact during SA.

Table 3.4: Nationalities of participants

Nationality	N	Percentage	Nationality	N	Percentage
Missing	14	6.3	Middle East	2	0.9
Afghanistan	1	0.5	Mongolia	1	0.5
Albanian	2	0.9	Morocco	1	0.5
Algeria	1	0.5	Oman	1	0.5
America	1	0.5	Pakistan	2	0.9
Angola	2	0.9	Portugal	1	0.5
British	2	0.9	Qatar	1	0.5
China	135	61.1	Russian Federation	2	0.9
Cote d'Ivoire	1	0.5	Saudi Arabia	5	2.3
Egypt	2	0.9	Chinese Taipei	8	3.6
Guinea	1	0.5	Thai	1	0.5
Hong Kong SAR	5	2.3	Turkey	2	0.9
Indonesia	2	0.9	Uganda	1	0.5
Japan	1	0.5	Ukraine	1	0.5
Kenya	1	0.5	Uzbekistan	2	0.9
Korea	11	5.0	Vietnam	1	0.5
Kuwait	1	0.5	Vietnam	1	0.5
Libya	1	0.5	Zambia	1	0.5

Malaysia	2	0.9	Zimbabwe	1	0.5
----------	---	-----	----------	---	-----

There were 51.9% male students and 47.2% female students, with 0.9% who preferred not to say. The participants' mean age was 21.07 years old (SD=3.59), with an average length of stay of 4.78 months at T1 (SD=8.19). 65.1% of participants had IELTS scores obtained within three months prior to T1 data collection and the median IELTS score was Overall Band 5.5 (SD=0.55). 39% of students had prior English-Medium-Instruction (EMI) experience and 36% of students had prior SA experience. For those with these two experiences, they had mean EMI instruction for four years (SD=3.77) and had studied abroad for an average of 1.6 years (SD=2.51). Regarding accommodation, 48.1% of the students lived by themselves, 25.5% lived with their co-national friends and 25.5% lived with international friends. Barely anyone lived with local British people.

Qualitative data collection recruited 30 interviewees in total. Only 18 participated in all three rounds of interview, among which 10 students did both the interview and the speaking test three times (Table 3.5). Interviewee attrition happened for various reasons: some students only volunteered to participate from T2 onwards when they trusted the researcher more; one dropped out from the course; some were not available due to sickness or personal reasons; some did not respond to the message for unknown reasons. There is a balanced ratio of interviewees with different PV scores at T1 (score<7: n=10, 8<score<12: n=10, score>13: n=9). Although the interviewee sample had slightly higher mean scores than the whole sample, the few low-scoring interviewees nonetheless represented those with lower English proficiency. Because more female students left their contact details than male students, there were 10 male and 20 female interviewees in the sample. Thus, more data were obtained from female interviewees.

Table 3.5: Interviewees profile

ID	Label	Gender	Nationality	Interview1	IELTS1	Interview2	IELTS2	Interview3	IELTS3
8110376	AC1	Male	Chinese	✓	✓				
8111469	AC2	Female	Chinese	✓	✓	✓	✓	✓	✓
8111535	AC3	Female	Chinese	✓	✓	✓	✓	✓	
8112165	AC4	Female	Chinese	✓		✓	✓	✓	✓
8110801	AC5	Male	Chinese	✓	✓	✓	✓		
8112422	AC6	Female	Chinese	✓	✓			✓	✓
8048684	AC7	Male	Chinese	✓	✓	✓	✓	✓	
8110668	AC8	Male	Chinese	✓					
8110967	AC9	Female	Chinese	✓	✓	✓	✓	✓	✓
8112153	AC10	Female	Chinese	✓	✓	✓	✓	✓	✓
8109478	AC11	Female	Chinese	✓	✓	✓	✓	✓	✓
8109861	AC12	Male	Chinese	✓	✓	✓	✓	✓	✓
8111724	AC13	Female	Chinese	✓		✓		✓	
8111871	AC14	Female	Chinese	✓	✓	✓		✓	
8112158	AC15	Male	Chinese	✓	✓	✓	✓		
8112764	AC16	Female	Chinese	✓		✓		✓	
8111808	AC17	Female	Chinese	✓	✓	✓	✓	✓	✓
8113041	AC18	Female	Chinese	✓	✓	✓	✓	✓	✓
8109688	BC1	Female	Chinese	✓	✓	✓	✓	✓	
8110193	BC2	Female	Chinese	✓	✓	✓	✓		
8111953	BC3	Female	Chinese	✓		✓			
8049135	BC4	Male	Chinese	✓		✓		✓	
8048132	BC5	Male	Chinese			✓	✓	✓	✓
8111264	BE1	Female	Egyptian			✓	✓	✓	✓
8111549	BR1	Male	Russian				✓		
8112205	BC6	Female	Chinese			✓	✓	✓	✓
8110416	BC7	Female	Chinese	✓	✓	✓	✓	✓	✓
8110442	BC8	Female	Chinese	✓	✓	✓	✓	✓	✓
8109948	BI1	Female	Ivorian	✓	✓	✓	✓	✓	✓
8112468	BK1	Male	Korean	✓		✓			

*Blue shaded cells were students who took all the interviews and IELTS speaking tests.

3.3.4 Participant attrition

The sample size was decided based on the high participant attrition rate in previous studies (e.g., 25% in Schmitt, Dörnyei, et al., 2004). To have at least 100 students completing three rounds of data collection for quantitative analysis, it was decided to recruit 250 students in the beginning. A total of 261 students from two foundation

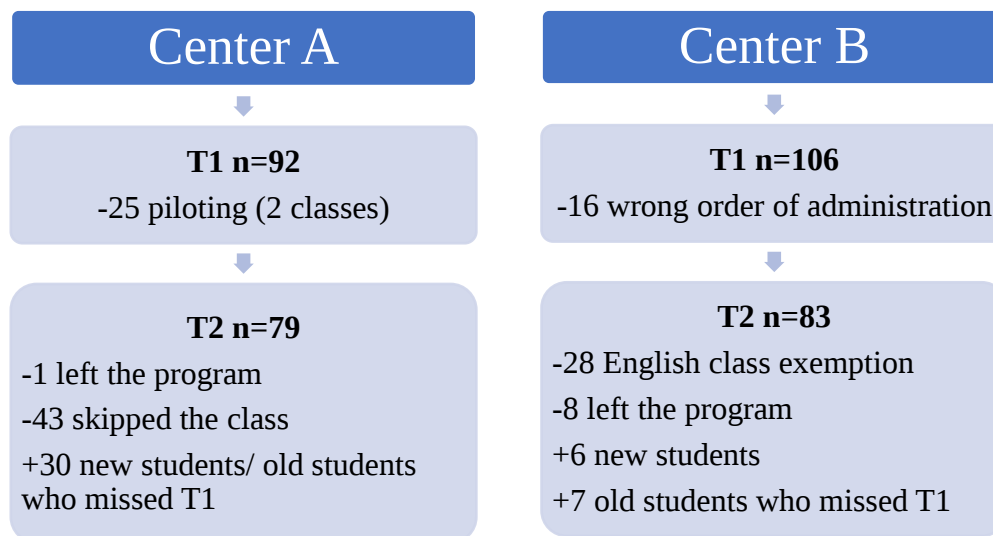
centers completed at least one instrument at one data collection point of the present study. At T1, T2, and T3, 221, 160, and 41 students participated respectively. The valid sample size for each quantitative instrument is shown in Table 3.6.

Table 3.6: Sample size for each quantitative instrument

Sample	Receptive PV test	Productive PV test	LCQ	SASNS
T1	151	158	176	121
T2	143	145	142	133
T3	19	19	40	23

At T1, some test scores had to be discarded due to unstandardized test paper distribution order and some SASNS data were discarded because some participants listed L1-speaking friends. Severe participant mortality occurred at T2 due to a few reasons (Figure 3.2). Center B had a program policy that students who passed the English exam in Term 1 were exempted from English classes in the next term. Although the researcher arranged a lunch break slot for these students who were exempted from English classes (about 31%) to complete T2 data collection and emailed the questionnaire links to these students, unfortunately, they did not show up or complete the questionnaire. At Center A some classes had poor attendance rates and some students were absent from the classes in which T1 or T2 data collection took place. It was later discovered that the foundation programs admit a few new students each term. The students who arrived in the UK at Term 2 (n=10) were marked as T1 participants in the analysis because their length of stay was equivalent to those who arrived before Term 1. The number at both centers plummeted at T3 because, during the UK COVID-19 lockdown period, classroom teaching was switched to online teaching. The course leaders could only send out emails with questionnaire links to participants instead of asking them to complete them in class. There were also a few participants who left the foundation program for personal reasons.

Figure 3.2. Reasons for participant attrition at T1 and T2



3.4 Research instruments

The current study employed a variety of instruments to answer the RQs. Considering the characteristics of the participants and the complexity of the study design, practicality was given priority and the length of each instrument was strictly controlled (see Section 3.5.1 for more details). In this section, quantitative measures will be introduced before qualitative measures in the order from dependent variables to independent variables.

3.4.1 Standardized PV tests

The present study employed both receptive and productive tests (see Appendix D) to measure gains in recognition and production of PVs during SA. As proposed by Henriksen (1999), receptive and productive vocabulary knowledge is not dichotomous but falls on a continuum. Because receptive knowledge usually precedes productive knowledge (Schmitt, 2010), tracking both types of mastery will capture knowledge development at either level of acquisition. The length, item selection, and other details of the test will be presented, followed by formats of the productive test and the receptive test, and a discussion of validity and reliability.

Tests on formulaic language normally have 30-60 items or more (Fernández & Schmitt, 2015; Gyllstad, 2005; Macis & Schmitt, 2017; Revier, 2009; Smith & Murphy, 2015). In studies with multiple instruments, it is not uncommon to use 20 items (Conroy, 2018; Schmitt, Dörnyei, et al., 2004) or less (Serrano et al., 2016). Due to the limited time for test administration and concern over *fatigue effect*, 20 target PVs were chosen and tested both receptively and productively at each time point (i.e., testing 40 items in total) and 10 minutes were allowed for each test based on various pilots. Polysemy was not tested in the current study due to inconsistent findings in a previous study (see Section 2.2.4) and the low English level of the participants.

As reviewed in Section 2.2.4, the target PV items were chosen from two well-designed tests (Garnier & Schmitt, 2016; Schmitt & Redwood, 2011). Although they were not substantially validated, they were designed with rigor with references to mega-native-speakers corpora: British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The PVs were chosen based on three criteria: 1) a stratified distribution of the corpus frequency in BNC (Liu, 2011) (i.e. seven PVs from the Top 1-50 range, seven PVs from the Top 51-100 range, and six PVs from the below Top 100 range); 2) a wide coverage of lexical verbs (i.e. 14 different verbs in 20 PVs) to make the test items more representative; 3) varied difficulty in a large-scale piloting test on foundation program students (Zhou, 2018), indicating that this test can differentiate students of varied PV competence. To account for possible huge gains in one academic year, two infrequent items (i.e. *pin down*, *take after*) from student coursebooks and grammar reference books in Schmitt and Redwood (2011) were included to avoid the possible *ceiling effect* of standardized tests (Table 3.7). The PV list by Liu (2011) was chosen because it is the most comprehensive high-frequency PV list published so far and it laid a foundation for much PV research in the last decade, as it has been regularly used and cited since publication (Garnier & Schmitt, 2015, 2016; Liu & Myers, 2018; X. Zhang & Wen, 2019). It should be noted that BNC has an imbalance proportion of corpus modality, with 90% written and 10% spoken. Thus, the overall corpus frequency ranking of PVs might be influenced by such imbalanced distribution. The semantic

opacity of formulaic language can be operationalized in a continuum or dichotomously. For example, Smith and Murphy (2015) categorized multi-word phrases into three types: transparent, semi-transparent, and non-transparent. For clarity, the current test borrowed the dichotomous judgment of items from Garnier and Schmitt (2016), while the semantic opacity of the items from Schmitt and Redwood (2011) was judged by the researcher (n=6).

Table 3.7: PV target items in the tests, ranked by BNC frequency

PV	Meaning	Meaning ranking ¹	Frequency in COCA ²	Frequency in BNC ³	Semantic opacity	Question Source ⁴
Go on	continue	2	1	1	Figurative	S&R
Go out	leave your house for a special reason	1	8	6	Literal	S&R
Come on	said as encouragement	1	25	13	Figurative	G&S
Come in	enter	1	14	14	Literal	G&S
Turn up	arrive, appear	3	70	38	Figurative	G&S
Get on	have a good relationship	/	127	39	Figurative	S&R
Break down	stop working	1	55	45	Figurative	G&S
Put out	extinguish	2	51	58	Figurative	G&S
Take back	return	1	72	60	Literal	G&S
Hold up	delay	3	38	61	Figurative	G&S
Cut off	remove	1	61	74	Literal	G&S
Run out	used completely	1	103	86	Figurative	G&S
Make out	see	1	104	89	Figurative	G&S
Get off	leave the bus	1	56	90	Literal	G&S
Reach out	stretch an arm to grab something	1	59	104	Literal	G&S
Hold back	contain, repress	1	118	112	Figurative	G&S
Stand out	distinguish itself by being better	1	111	113	Figurative	G&S
Come off	be detached, separate from	1	132	136	Literal	S&R

Pin down	make someone decide	1	N/A	N/A	Figurative	S&R
Take after	similar to, be like	1	N/A	N/A	Figurative	S&R

¹ Corpus ranking in Garnier and Schmitt (2015)

^{2,3} Corpus frequency ranking from Liu (2011)

⁴ S&R = Schmitt & Redwood (2011); G&S = Garnier & Schmitt (2016)

N/A means this PV is not in the most common 150 PV list

When adapting the test, to ensure the questions fit young students from diverse cultural backgrounds and with low English proficiency, daily-life text (e.g., Our car broke down yesterday) was preferred over politically-related text (e.g., Following a disagreement, the government broke off diplomatic relations with China.). Neutral text (e.g., The fire has finally been put out.) replaced text related to crime or police (e.g., Police have put out a warning about thieves in the area.). Difficult words were replaced by more common words (e.g., reach out to the *jar* was changed to reach out to the *glass*). The English definitions were not changed to maintain test validity. The most frequent meanings of the target PVs in COCA were chosen for most PVs. The five exceptions of choosing a lower-ranking meaning were either because the context of the original prompt was not ideal (i.e., related to politics or crimes) or students might be more familiar with the tested meanings based on piloting (e.g., get on well with friends; sb. turn up in class).

In terms of administration format, off-line pen-and-paper tests were employed in previous studies (Fernández & Schmitt, 2015; Garnier & Schmitt, 2016; Macis & Schmitt, 2017). Though the current study has a large sample that seems to favor online testing, after piloting both the online test and the offline test, a pen-and-paper format was adopted to prevent students from cheating or getting distracted by other mobile phone applications. Example questions were provided in both tests to help learners to understand the instructions. Two marking protocols were trialed for all the test papers. One method gives one point to the correct combination of the verb and the particle, while the other method gives one point each to a correct verb and a correct particle. The results in gains analysis for both marking protocols were the same. However, since

getting half of the PV right could happen by chance, the former method was adopted for analysis. In the current study, minor spelling mistakes were accepted, following the marking protocol in Schmitt and Redwood (2011). For example, *came* and *coming* were accepted as the correct alternative to *come*.

3.4.1.1 Productive test

This cloze productive test measures the form-meaning mapping of PVs. Among productive tests on formulaic language, cloze tests might be the most widely adopted format for their clear and user-friendly design (Fernández & Schmitt, 2015; Garnier & Schmitt, 2016; Macis & Schmitt, 2017; Schmitt & Redwood, 2011). They require learners to spell the target PVs correctly based on definitions given in brackets, the scenario described in the sentence, and the initial letters provided in the sentences (see Table 3.8). In this example, *come on* should be the correct answer. When sorting the items, high-frequency PVs were put at the beginning of the test to boost students' confidence. The difficult questions were put at the end, to minimize possible score loss if students get discouraged.

Table 3.8 Sample item in the productive PV test

1. C ____ o ____, don't be shy and tell us your story! (<i>said as encouragement</i>)

3.4.1.2 Receptive test

The present study deployed an innovative multiple-choice test format to motivate learners, considering they were going to be tested twice on the same items in a short period of time. The present study opted for a modification of a Multi-word Phrases Test (MPT) by Smith and Murphy (2015), which was adapted from a collocation test called CONTRIX (Revier, 2009). The original CONTRIX provides a 3×3 matrix of verbs, articles, and nouns for learners to choose from. Following the MPT, the current test retains a 2×3 matrix of verbs and particles. Test-takers need to choose a lexical verb from the left column and a particle from the right column (Table 3.9), to form a correct PV that fits into the scenario. The merit of the MPT format is that learners need to recall

the *whole* phrase, instead of only *part* of the phrase as in many receptive tests, which can minimize the *guessing effect*. The chance of guessing the right meaning without any partial knowledge of the PV is 1/9 (11%).

Table 3.9 Sample item in the receptive PV test

31. Our car _____ yesterday.	
Broke	Along
Brought	Down
Kept	Off

The PV items were the same as the productive test, but the order was rearranged. After comparing the 3×2 matrix (i.e., 3 columns and 2 rows) and the 2×3 matrix (i.e., 2 columns and 3 rows) in piloting, the latter format was chosen in favor of a clear layout for test-takers. In terms of distractors, since in piloting some students could remember the initials of the PVs from the productive test they took earlier, at least two verbs or two particles starting with the same initial were designed for each question in the final test (e.g., *broke* and *brought* in the example). For each item, at least one distractor looked like a plausible answer. Due to the flexible nature of PVs, the distractors could make up multiple PVs (e.g., *bring along*, *bring down* in the example). Thus, the distractors could not control the number of possible PV combinations. All the verbs and particles were listed in alphabetical order to avoid giving any extra clue to test-takers.

3.4.1.3 Validity and reliability

There is a lack of literature on the validity and reliability of existing PV tests, since limited attention was paid to PV tests compared with one-word vocabulary tests. The validity of the current PV tests can be examined from a few perspectives. Firstly, Garnier and Schmitt (2016) and Schmitt and Redwood (2011) piloted the tests on native speakers of English and modified the prompts accordingly. Second, criterion-related validity, which has been questioned (Fitzpatrick, 2007), could not be examined in the present study, because of the absence of other previously validated PV tests for correlation analysis. As for the content aspect of construct validity, the current tests had

a clear boundary of measuring the form-meaning relationships of verb-particle constructions. In terms of external validity, the PVs with the highest means and the lowest means in the productive PV test were almost the same for mixed-L1 students in the UK (S. Zhou, 2018) and Chilean English-major students (Garnier & Schmitt, 2016), suggesting this test can be applied to populations from different countries and yield relatively consistent outcomes. Future research is needed to conduct item analysis, factor analysis, interviews with participants, and so on, to better validate the test.

Regarding reliability, Garnier (2016) reported a Cronbach's α of .95 in her 100-item productive test. A 45-item productive pilot test ($n=118$) and a 30-item pilot test ($n=46$) by the researcher on international foundation students in the UK reached good internal reliability of $\alpha = .933$ and $\alpha = .835$ respectively. With regards to the receptive test, the original MPT has high reliability (Cronbach's $\alpha = .85$) and good test-retest reliability ($r = .87$, $p < .001$) (Smith & Murphy, 2015). The final 20-item test reached internal reliability ranging between .73 and .82 (Table 3.10). Further, in piloting, students' performances on the same receptive test in the online setting and the offline setting were similar: $t(28) = .639$, $p = .528$, showing strong reliability across testing mediums. Lastly, three rounds of data collection were set in weeks when students had no exams, to increase students' willingness to participate. The tests were completed in the classroom under the supervision of the researcher/teachers, to control test conditions and strengthen internal reliability. Overall, although statistical validation of the test was not carried out due to limited participant resources, based on various piloting and existing literature, the adapted tests did not spot any major flaw in validity and reliability.

Table 3.10 Internal consistency of receptive and productive PV tests

PV tests	P1	R1	P2	R2	P3	R3
Cronbach's α	.782	.728	.763	.797	.770	.817

Note: P=productive test, R=receptive test, 1=T1, 2=T2, 3=T3

3.4.2 the Vocabulary Levels Test

The present study included a receptive updated VLT (Webb et al., 2017) at T1 and T2 data collection, to understand the impact of initial vocabulary knowledge on PV acquisition. The VLT was first developed by Nation (1990) and updated by various researchers (Schmitt et al., 2001), and is widely adopted internationally as a diagnostic test or placement test (Kremmel & Schmitt, 2017). Rather than giving an estimate of the learners' vocabulary size by specifying how many words they actually know, it provides a profile of learners' receptive vocabulary commands at different frequency bands (Schmitt et al., 2001). For example, a learner might know some low-frequency words in the 5k level, but it does not mean he or she knows 5000 words in total. The advantage of the VLT is its strong practicality to distinguish students' vocabulary levels with a minimum amount of reading and within limited testing time. The VLT has strong validity and reliability, as demonstrated by Schmitt et al. (2001). The updated VLT was also validated by Webb et al. (2017) from five aspects of construct validity and it has a Rasch item reliability of .96.

The decision to include the VLT as a baseline measure was enlightened by Schmitt, Dörnyei, et al. (2004) who tracked the formulaic sequence development of a similar population to the current sample over a 3-month period, in which VLT 3k and VLT 5k were administered at the beginning of data collection. The researcher piloted VLT 2k to 5k on 17 learners with IELTS scores from Overall Band 5.0 to 7.5 and some high achievers (IELTS 6.5 or above) obtained near full marks in VLT 2k and 3k. Therefore, the VLT 5k was chosen to distinguish the vocabulary size of participants of varied proficiency sufficiently. Admittedly, only administering the VLT 5k cannot assess learners' vocabulary breadth of other frequency bands. Since Schmitt et al. (2001) increased the number of items per level from 18 to 30 to improve reliability, it is believed that using all the items in one level has higher reliability than testing a handful of items from two levels.

The receptive updated VLT (Webb et al., 2017) was chosen in the present study because

it drew on more recent corpus data in Nation (2012) and it tested the knowledge of three target words in one item with high efficiency (Table 3.11). Although this design might affect the item independence, Webb et al. (2017) chose words with very different meanings to avoid such influence on test-takers. Participants were given five minutes to answer 10 items that tested knowledge of 30 words. The VLT was administered at T2 as well, to prevent students from realizing that PV knowledge was the dependent variable of central interest in the present study. This test has two equivalent versions (Webb et al., 2017) which can be used in multiple testing in longitudinal studies, so the researcher used Version A at T1 and Version B at T2.

Table 3.11 Example VLT item

	game	island	mouth	movie	song	yard
land with water around it		✓				
part of your body used for eating and talking			✓			
piece of music					✓	

The productive VLT was not used because asking participants to do two productive tests consecutively was deemed to cause fatigue. The New VLT (McLean & Kramer, 2015) was not adopted, although it is also based on Nation's 2012 BNC/COCA word lists. Its multiple-choice format has the pros of good *item independence* and easiness for online administration and the cons of a large item number and long test length. Considering the offline testing format, the updated VLT was preferred over the New VLT. Another validated vocabulary size test, the Eurocenters Vocabulary Size Test (Meara, 1990), was not adopted because it measures learners' self-perception of vocabulary knowledge, rather than actual knowledge (Briggs, 2015a), which might undergo score inflation.

3.4.3 Questionnaires

The questionnaires on Qualtrics consist of two parts: the LCQ and a demographic questionnaire (only for T1). The SASNS was run on another website. The questionnaire links were accessed from QR codes by the participants on their phones and shortened URLs were also provided in case they could not scan the codes.

3.4.3.1 Language Contact Questionnaire

The present study developed the LCQ (Appendix E), based on the Language Contact Profile (LCP: Freed et al., 2004) and the Language Engagement Questionnaire (LEQ: McManus et al., 2014). The LCP and the LEQ both had some redundancy issues as discussed in Section 2.3.1.2. The present LCQ aims to adapt existing instruments for specific research purpose while recognizing the limitations of self-report data.

Figure 3.3 Sample item in the LCQ

Thank you for coming back! It has been three months since we did this survey last time. In this language contact questionnaire, we would like to know **how often you use English** in your spare time **in the UK**.

1. Read English books, newspaper, or magazines

never

rarely

sometimes

often

very regularly

all the time

Since the present study aims to analyze the holistic language contact pattern of a large number of learners, fewer items shall suffice (Baten, 2020). Thus, the LCQ includes 16 L2-related items, to present a global view of learners' L2 use over an extended period of stay. The order of the items follows the original LCP, from receptive activities to productive activities. A six-point frequency scale was adopted, being *never*, *rarely*, *sometimes*, *often*, *very regularly*, *all the time* (Figure 3.3), referring to the five-level scales in the LEQ and the one in Serrano et al. (2016). The researcher believed that these scales are less quantifiable but more reliable (Serrano et al., 2016). Since the data

collection always took place during term time, the LCQ mainly reflects the term time L2 use of the participants. The final internal reliability of the LCQ, measured by Cronbach's α , was .846, .847, and .880 for T1, T2, and T3 respectively.

Some other minor changes were made to the LCQ. For example, following Briggs (2015b), the word “*try*” was deleted from each original LCP statement to avoid capturing strategic behavior as opposed to actual contact. Secondly, similar to the LCP, the LCQ retains a focus on L2-only activities, with a supplement of the two most common L1-using activities as determined by results in a pilot study (S. Zhou, 2018) as a comparison (i.e., using social media and speaking with co-national friends). The emphasis was placed on L2 activities because L1-use are not likely to contribute to PV acquisition. To highlight that these two questions were different from the rest, the text “in your first language” was made bold. Lastly, the original LCP did not foresee the proliferation of telecommunications technology (Dewey, 2017). Thus, items relating to the Internet and instant messaging were added, as per the LEQ. After interviewing some students in piloting, “playing computer games”, which did not appear in either the LCP or the LEQ, was added to the LCQ.

The LCQ was administered on Qualtrics in both Chinese and English, considering the high percentage of Chinese students in the learner cohort. In previous studies, it is commonplace to distribute language contact instruments in the L1 of the participants (Freed, Dewey, et al., 2004; McManus et al., 2014). In this study, based on descriptive statistics of the VLT and PV tests, non-Chinese students had a higher English level than Chinese students. Therefore, offering the Chinese version of the questionnaire was an attempt to reach an equal level of comprehension for all participants and enhance the validity of the results. Although the Chinese version inevitably poses a threat to the reliability of results, the researcher made sure to only use simple English in the LCQ. The online format was chosen considering the large sample size, convenience of data collection and data storage, and paper-saving purpose.

3.4.3.2 Demographic questionnaire

The demographic questionnaire collected the following information: types of foundation program (pre-bachelor or pre-masters), age, gender, nationality, and accommodation type in the UK, to illustrate a profile of the participants (see Section 3.3.3 for sample characteristics). A unique factor affecting PV acquisition is the potential L1 transfer (Boers, 2019). Therefore, collecting the nationalities of participants is useful. On top of that, students were asked what IELTS scores they had, whether they had received EMI previously or whether they had studied abroad before, and the duration of these two forms of education. These pieces of information were obtained for two reasons. Firstly, empirical studies have confirmed a positive correlation between vocabulary knowledge and proficiency test performance, such as IELTS (Qian & Lin, 2019). Since IELTS scores might be outdated, only IELTS scores obtained three months before T1 were collected. Secondly, given the positive association between frequent encounters and uptake (Durrant & Schmitt, 2010; Siyanova & Schmitt, 2008), it is likely that EMI experience and SA experience can aid the intentional or incidental acquisition of PVs (Boers, 2019).

Eight motivation questions were asked in the same demographic questionnaire at T1 as a baseline measure of a possible confounder influencing the dependent variable and independent variables in this study. Since the emergence of the L2 Motivational Self System (L2MSS, Dörnyei, 2009), there have been multiple L2MSS surveys developed for Chinese learners (T. Taguchi et al., 2009; You & Dörnyei, 2016). The researcher adopted questions from previously validated questionnaires (Csizér & Kormos, 2009; Iwaniec, 2014) with multi-item scales and simple language that suits less proficient English learners. As for the answer format, the five-point Likert scale in Iwaniec (2014) is favored over the six-point Likert scale as in You and Dörnyei (2016), to allow for neutral or indifferent responses from participants. In the current study, the piloted 12-item questionnaire had a Cronbach's α of .824 and the final 8-item version's Cronbach's α was .781 (N=174).

3.4.3.3 The Study Abroad Social Network Survey

Social networks research can shed light on the level of integration with host-nationals in the local community (Dewey, Belnap, et al., 2013), so collecting both language contact and social networking data has become an increasingly popular practice in SA literature, to understand the out-of-class experiences of SA participants (Baten, 2020; Strobl & Baten, 2021).

The exponential growth in social network studies gave rise to plenty of digital methodological innovations in different disciplines (Hogan et al., 2020; Stark & Krosnick, 2017). In the present study, the SASNS was developed based on the Study Abroad Social Interaction Questionnaire (SASIQ: Dewey et al., 2013), the Social Network Questionnaire (SNQ: McManus et al., 2014), the personal network questionnaire (Xie, 2017), and the Social Network Strength Scale (SNSS: Kennedy Terry, 2017), with technical support of an Oxford Computer Science alumnus. The [SASNS](#) is run on Amazon Web Service (an online platform that collects and stores information) and can be opened as a webpage, meaning that respondents can avoid the hassle of installing an application on their mobile device (e.g. the Network Canvas, Hogan et al., 2016). The software was tested multiple times from collecting data to exporting data, to improve software bugs and ensure that the instructions for each step were clear to the respondents.

To date, there is little appraisal of the validity and reliability of social network questionnaires in SA research. In terms of face validity, these instruments have a clear boundary of what they intend to measure. The SASIQ and the SNSS only concern face-to-face networking with L2 native speaker interlocutors, whereas the SNQ provides a comprehensive picture of L1, L2, and even L3-speaking friends and categorized social networks in different social contexts. The lack of convergent validity checks and alternate-form reliability checks should be addressed in future SA studies. A threat to the external validity of existing questionnaires is their small sample sizes, being 17 and 29 respectively for the SNSS and the SNQ. The SASIQ was adopted in a series of SA projects whose respondents added up to 118 in six countries, with a reported test-retest

reliability from .75 to .95 (Dewey et al., 2014). Xie (2017) contended that visualized network representations could enhance the validity of the network data by facilitating survey engagement, which is adopted by the SASNS in the density measure.

Figure 3.4 Structure of the SASNS and reference instruments for each part

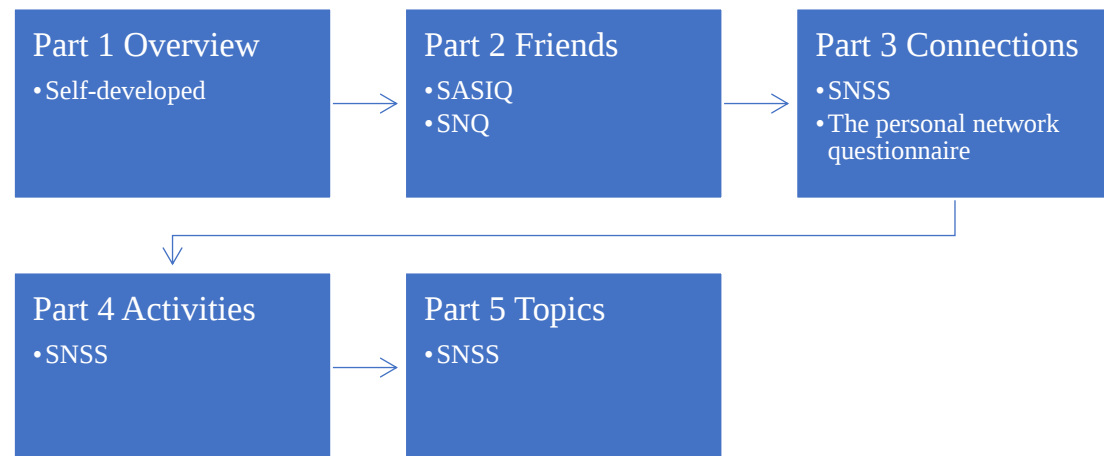


Figure 3.4 outlines the components of SASNS and the existing reference instruments of each component. The first page is an overview that collects the size of students' personal networks. The pages of Part 2 depend on how many friends the students have. One page is dedicated to one friend and up to five friends can be listed. Part 3 shows the connections between the respondents' friends. Part 4 and Part 5 concerns what activities and what topics the students do or talk about with their friends. On the last page, student ID was collected as a personal identifier. The design of each part will be elaborated below.

Part 1 of SASNS (Figure 3.5a) forms an overall impression of students' social networks. Three pieces of information were collected on this page, being the number of L1-speaking friends, the number of English-speaking friends, and the number of their English-speaking social circles. They correspond to three variables in further analysis, namely, the sizes of L1 and L2 social networks and the dispersion of L2 social networks (see Table 3.14 in Section 3.6.2.). It should be emphasized that SASNS only focuses on L2-speaking friends. Including the L1-speaking network size is just to help identify the proportion of L1 and L2 friends.

Figure 3.5a Part 1 of SASNS

Social Network Survey 1.3

Select language

English ▼

This survey is about your English-speaking friends in the UK. Please answer the question truthfully and your privacy will be protected.

How many friends do you have in the UK that speaks your first language?

Enter number

Now think about your **English-speaking** friends who meet ALL of the following conditions:

1. You have met each other in person. You know each other's name.
2. You have each other's contact detail (e.g. phone number/Facebook/WhatsApp).
3. They are not from your home country.

How many **English-speaking** friends (from the UK or other countries) do you have in the UK?

Enter number

How many **English-speaking** social circles (friendship groups) do you have? (e.g. classmates, flatmates)

Enter number

Continue

While the definition of L1-speaking friends was loose, the criteria for counting English-speaking friends are: 1) they are not from the same country. 2) they know each other's names and have met in person. 3) they have at least one contact detail of their friend (e.g., phone number, WhatsApp, Facebook). To remind the respondents that “English-speaking” was not limited to native speakers, the explanation “from the UK or other countries” was included. These inclusion criteria were looser compared with the SNSS and the SASIQ. The SNSS has a high inclusion threshold in that only L2 native-speaking friends with whom learners speak for at least 30 minutes per week can be listed (Kennedy Terry, 2017), which does not account for the L2 input from non-native L2 users. One of the inclusion criteria for SASIQ is “you know them well enough to

have spent at least some time socializing with them”. Foundation year students tend to have superficial cross-cultural friendships and a low frequency of L2 socialization, so having a high threshold will limit the number of friends they can list. As for “social circles”, both the definition (i.e., friendship groups) and two examples (e.g., classmates, flatmates) were given.

In Part 2, respondents were tasked with providing up to five L2 friends’ detailed profiles. The upper limit of five was set considering the motivation and time commitment of the participants. Initially, the lower limit and upper limit were set as 3 and 10 friends, following existing literature (Isabelli-García, 2006; Xie, 2017; Y. Zhou & Todman, 2009). After piloting on foundation students in the UK (n=21), the limit was lowered to one and five, because most pilot participants just bothered to list the minimum number of three friends to five friends. With the researcher’s encouragement, two pilot participants wrote 10 friends, but their lists included four and seven co-national friends respectively. Thus, setting the limit to 10 was deemed unnecessary, given that the size of the L2 network was already collected in the overview. The participants in Xie (2017) were from one of the top two universities in the UK and they were highly proficient in English. Considering distinct sample features, the modification could be justified. After T1 data collection, it was found that some respondents did not complete the SASNS because they did not have any foreign friends who met the inclusion criteria. Therefore, at T2 and T3, participants were allowed to list zero friends if that was the case. L1-speaking friends’ profiles were not collected because having more L1-speaking friends is not likely to contribute directly to PV development.

In Part 2, respondents were tasked with providing detailed information about their friends, ranging from their gender, nationality, the occasion of initial encounter, to their English proficiency, the frequency of online and offline contact, their relationship with the respondents (most of these dimensions were adapted from the SASIQ or the SNQ). These data were collected because they either help us to understand the typical characteristics of the English-speaking friends of SA students or provide hints of possible linguistic impact on respondents’ PV development. Four social network

variables were yielded from this part, being friend's English proficiency, online contact frequency, face-to-face contact frequency, and intensity of friendship. A question in the SASIQ asking the average monthly contact hours with each friend was discarded, because it might suffer from *recall bias*. On one hand, it is hard for students to estimate the total hours if they meet their friends very often. On the other hand, as a *prestige bias*, respondents might consciously or unconsciously overestimate the time that they spend with L2 friends to look more sociable.

Figure 3.5b Part 2 of SASNS

Friends

Please give the information of up to 5 closest English-speaking friends you have.

Name (can be first name or nickname)

Gender

How did you meet

How good is your relationship

Nationality

Your friend's English proficiency

Frequency of face-to-face contact **in English outside of class**

Frequency of online contact **in English**

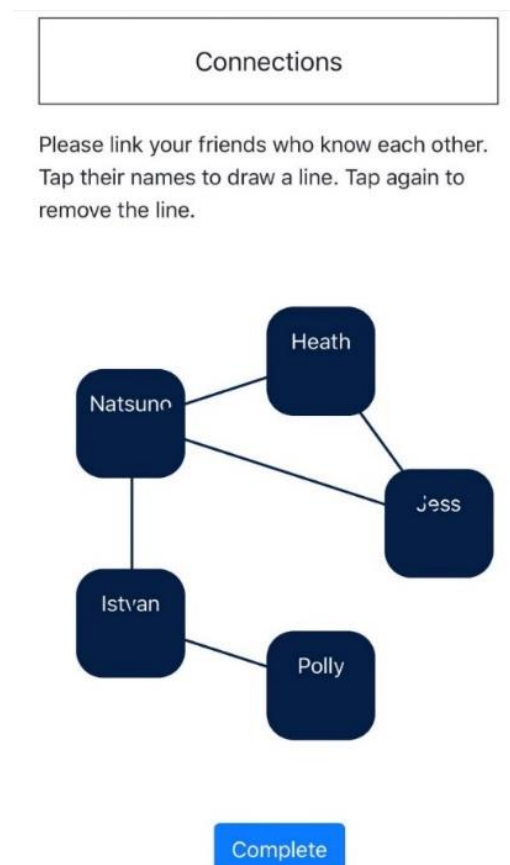
<
Friend 1 of 1
>

Delete friend
Add friend
Complete

Drop-down menus were adapted from the SASIQ and were provided for each question, to facilitate ease and speed of responses and data coding (Figure 3.5b). For instance, the open-ended question asking respondents how they made “native-speaking” friends in SASIQ was adapted to this format. Except for nationality which the respondents

might not remember, all other blanks were compulsory to fill. For example, the English proficiency of their contacts was categorized into: *Lower than me/Similar to me/Better than me/Native*. Studies showed that speaking English with fluent non-native L2 speakers can also lead to proficiency gains (Magnan & Back, 2007). Only distinguishing “native” and “non-native” speakers of English cannot capture the different proficiency levels among non-native speakers. Therefore, the researcher devised this four-level scale, which lays emphasis on relative English competence, not absolute competence. The intensity of friendship is graded on a 4-point Likert scale from *very good* to *not so close*, instead of the 10-point Likert scale as per the SASIQ, for succinct coding. The contact frequency is broken down into face-to-face contact and online contact on four scales, being *rarely*, *every month*, *every week*, and *every day*, similar to the scale in the SASIQ: *very often*, *often*, *sometimes*, *rarely*, *never*. Since technology has exerted a profound influence on the SA experience of sojourners, online contact needs to be considered (McManus et al., 2014).

Figure 3.5c Part 3 of SASNS



In Part 3, all the friends that respondents had named appear in a pentagon shape on the screen (Figure 3.5c) in which respondents can physically link people who know each other by tapping their names consecutively. This interface was deemed interesting and straightforward for the participants. The number of lines drawn in this part plus the number of listed friends yielded the variable of network density (i.e., the degree of connections between people within one's network).

Figure 3.5d Part 4 of SASNS

Activities

Overall, have you done any of the following activities with your English-speaking friends in the **last month**? If yes, please put the number of times for each activity.

Share a meal

Enter number

Have coffee

Enter number

Go to gym or play a sport

Enter number

Play board games or cards

Enter number

Go shopping

Enter number

Go to bars/clubs

Enter number

Figure 3.5e Part 5 of SASNS

Topics

Overall, what topics have you discussed with your English-speaking friends out of class in the last month?

- ☐ Home country vs. British culture
- ☐ Celebrities
- ☐ Clothing and fashion
- ☐ Gossip
- ☐ Hobbies
- ☐ Holiday and travel
- ☐ Music
- ☐ Politics
- ☐ Sports
- ☐ Television and movies
- ☐ Problems I am having (e.g. being homesick)

Finally, Part 4 and Part 5 are the multiplexity measures (i.e. the richness of the learner's English interactions, Kennedy Terry, 2017) of the SASNS (Figure 3.5d, Figure 3.5e). Elements from the SNSS are integrated into the SASNS, in which students need to count, as a whole, what activities they have done with their friends in English in the past month, and check the *breadth* (i.e., the number) of topics they have discussed in English. Activities and topics in the SNSS were adapted, by removing the rarely

mentioned item (e.g., technology) and adding new items based on piloting interviews (e.g., gossip). Finally, 10 activity items and 13 topic items were chosen in the final version. The multiplexity variables might have a unique value for investigating vocabulary development because more varied input increases the chance of learners encountering and noticing unfamiliar PVs (Long, 1983). Open-ended space was left at the end, enabling respondents to add any unlisted activity or topic in the SASNS.

3.4.4 IELTS speaking test

To assess the command of productive PVs of the sample, the IELTS speaking test (Appendix F) was conducted at every time point with the researcher as the conversation guide/interlocutor. The IELTS speaking test has three parts (i.e., short warm-up conversation, individual monologue, and deep discussion), with each lasting about 4-5 minutes (Karim & Haq, 2014).

This test was chosen as the spoken productive PV knowledge elicitation measure because an IELTS score is a prerequisite for most international students to pursue a degree in British universities, and the sample is supposed to be familiar with this test format, which can reduce their anxiety and increase external validity. Also, the IELTS speaking test can efficiently elicit spoken discourse about varied topics from learners in a limited time (11-14 minutes). The test was not ‘marked’ because it serves in the present study merely as a PV elicitation tool. The PVs were analyzed from a few aspects, which will be presented in Section 3.6.1.2. A huge amount of work went into validating the IELTS by academics and the test organizing authorities and some validations were dedicated to the speaking test (Elder & Wigglesworth, 2006; O’Sullivan & Lu, 2006; Zahedkazemi, 2015). It was shown that the IELTS speaking test not only had good construct validity when being compared with other standardized tests but also reached consistent results across different settings and learners of different proficiency.

Existing research shows that one of the key lexical differences between more proficient and less proficient learners in the IELTS speaking test is their use of formulaic language, with more proficient learners using more high-frequency formulaic expressions (Read

& Nation, 2006). Thus, it is possible that the participants would witness growth in their spoken PV production over one academic year, alongside their overall English improvement. Since it is common for Chinese learners (and other learners whose L1 does not include PVs) to avoid using PVs (Liao & Fukuya, 2004), different topics were tested in piloting. The selected topics involve daily life habits/experiences, rather than descriptions of an object or a person, to increase the opportunity for using PVs. The same categories of topics were used for each data collection point as a means of maintaining internal reliability (Table 3.12). For example, travel-related topics were included in all time points for consistency. The reason for using it again at T2 is that many students traveled in the UK or Europe during the Christmas holiday, so this topic might help them to produce longer utterances based on their recent trips.

Table 3.12 IELTS speaking test topics

	Part 1	Part 2	Part 3
T1	Studies, free time	An interesting journey	Healthy lifestyle
T2	Hometown, hobbies	An interesting journey	Leisure activities
T3	Friends, evening routines	An interesting film	Travel, transportation

3.4.5 Semi-structured interview

The researcher conducted semi-structured interviews (see Appendix G for the interview schedule) of the subsample at each time point. The questions focused on three topics, being their informal language contact, their social networks, and their vocabulary learning experiences. The length of interviews was controlled within 15 minutes, following the IELTS speaking test. To minimize the influence of participating in the current study on their learning behavior, only in T3 participants were asked specific PV learning questions, while in T1 and T2 they were asked questions about general vocabulary and phrases. To avoid an impact on students' confidence, T1 and T2 interviews did not reveal students' test scores. At T3, the interviewees were told their LCQ scores and PV test scores at T1 and T2 and were asked about the possible reasons for the stagnation, gains, or attrition of PV knowledge, to understand their perspectives on the change. Some interview questions in Wright and Schartner (2013) for Chinese

students were adopted, as they addressed the social interaction of sojourners. Since the researcher only got to meet these participants once per term, each interview started with a warm-up question. All the interview questions were reviewed by an L1 speaker of English prior to use.

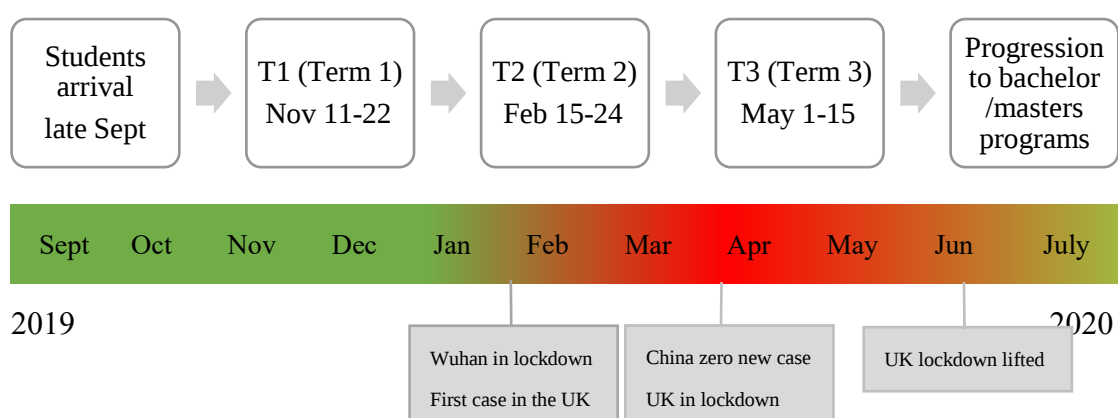
To enhance the validity of the interview, the researcher reduced the power distance, reassured their anonymity, and built rapport with the interviewees. The researcher always explained the structure of the interview at the beginning of it, expressed empathy when needed, and consciously accumulated knowledge of their foundation centers (Cohen et al., 2017). At T3, member checks were performed because of a concern over overseas video call quality. The interviewees confirmed that the transcript was an honest, accurate, and comprehensive record of their interview. In the present study, considering that social network relationships might involve some sensitive issues, students were given the autonomy to choose their preferred language for interviews where linguistically possible (Chinese/English), to minimize language barriers. It should be acknowledged that doing interviews in two languages might lead to concerns over validity and reliability. To be specific, dependability (the equivalent of reliability in qualitative research) (Cohen et al., 2017) is affected because learners tend to be more eloquent in their L1. The tradeoff is worth it because the majority of the respondents are Chinese and allowing them to take the interview in their L1 is beneficial to question comprehension and easier opinion expression. Finally, the details of the informants, research settings, and methods are provided in the thesis to improve external reliability.

3.5 Procedure

This study collected data from the same group of participants three times in one academic year (Figure 3.6). In the planning stage, the researcher visited the Head of English of the education group to pin down two target centers for participant recruitment and a rough timeline for data collection (avoiding final exams and holidays). Before each time point, the researcher held meetings with the course leaders to discuss

details of data collection (e.g., classrooms and time slots). Afterward, the researcher spent one week in each city to carry out quantitative and qualitative data collection and ensured that data were collected within a two-week window for each round. The researcher carried out all data collection by herself except the quantitative data at Center B at T1, due to safeguarding concerns from the Head of the center. At T3, due to the lockdown in the UK, both the course leader meetings and the data collection were completed online.

Figure 3.6 Study timeline



Note: The green part of the timeline means no COVID cases in the UK, and the red part means a high number of new COVID cases.

3.5.1 Study piloting

The present study involves a number of instruments, which requires extensive piloting. The researcher first recruited mixed-L1 SA students with B1-B2 levels from three private language schools (n=32) and one foundation program (n=42) in Oxford to test the reliability, validity, and administration of all the instruments. These participants resembled the target sample in their English proficiency, nationalities, and an extended stay in the host country for social network establishment. However, the proportion of Chinese students in those samples was small. Therefore, to understand Chinese students' vocabulary command, two first-year classes from a polytechnic in China (n=54) were recruited to complete the PV tests, and some Chinese learners of varied English proficiency (n=17) were recruited for the VLT level selection. The language contact

questionnaires were not distributed to these Chinese participants as they did not study abroad. Piloting started in August 2019 and finished in October 2019.

Piloting influenced the study design in a few ways. Overall, the pilot participants had limited patience to complete more than two instruments and they did not want to spend time participating in academic research after class. Since the present study utilizes four quantitative instruments, the formal data collection follows the *principle of parsimony* (Cohen et al., 2017). That is to say, each data collection should be controlled within 30-40 minutes (unlike one hour in the original plan) and has to take place in class. It was found that some participants could remember the PV initials from the productive test and used this memory as an aid in the subsequent receptive test: as such, it was decided that there needed to be a long interval between the two PV tests, and therefore the questionnaires were administered between the PV tests.

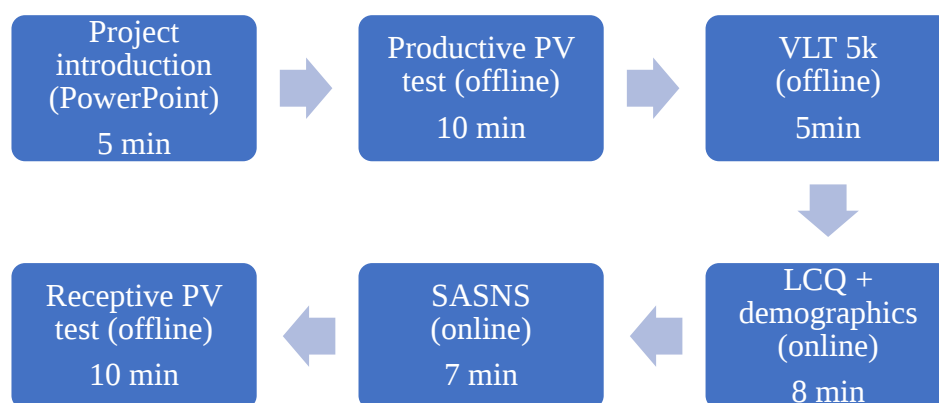
With regards to the PV tests, piloting was done to check the normal distribution, design of distractors (avoid self-evident answers), testing methods (online/offline), difficulty (mean score of each PV), and reliability (Cronbach's α). There was an open-ended question at the end of the test for students to leave feedback. Many students found the test hard, but some found the receptive format interesting. The PV tests were finalized after a few rounds of revisions. The non-significant difference ($p=.528$) between the test scores from the online and offline formats of the receptive test on learners of the same level enabled the researcher to choose either format. The reason for choosing the offline test came down to the purpose of making students take it seriously (i.e., it more closely resembled a typical, formal university exam format). As for speech samples entailing PVs, the ability of the IELTS speaking test to elicit PVs was compared with the preponderance of PVs elicited via the frog story (Berman & Slobin, 1994), which is another option for eliciting L2 speech samples. However, the frog story was deemed a little childish by the pilot participants ($n=5$) and did not interest them.

As regards the language contact measures, the questionnaire was checked for ease of understanding of the rubrics, items, and response scales. The LCQ was perceived easier

to complete than the SASNS by the pilot sample because the receptive format demanded fewer cognitive resources than open-ended questions. It was found that some students' phones could not scan QR codes on the screen, so shortened URL links and QR codes were both printed on paper. Semi-structured interviews were piloted on five Chinese foundation program students and considerable challenges were reported by them in terms of study, survival, and L2 socialization. Considering the English proficiency of the interviewees, L1 use was allowed for efficient communication.

3.5.2 Time 1 data collection

Figure 3.7a Procedure of T1 data collection



The researcher visited each class in the pre-arranged slots during their English classes (either at the beginning or the end of the class). The order of instrument administration in each data collection was identical. T1 quantitative data collection lasted for 45 minutes, with 5 minutes for presentation and 40 minutes for data collection (Figure 3.7a). A PowerPoint presentation was given to introduce the researcher's background, the purpose of the study, and the steps and rules of data collection. Straight after the presentation, data collection took place in class and informed consent was obtained from the participants.

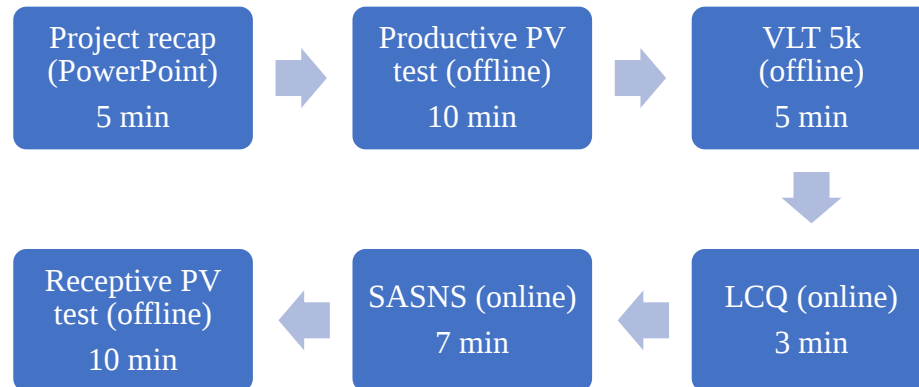
Firstly, the researcher distributed a test paper including the productive PV test on one side and the VLT 5k on the flip side (Appendix D). Participants completed these tests under supervision in that order. An oral reminder was given when there were five

minutes left for the PV test and two minutes left for the VLT (as it marks half of the test time had passed). Then the researcher collected the first paper and distributed the second test paper, to prevent cross-referencing. Participants were asked to take out their phones and scan the QR codes on the paper (Appendix D). The first QR code directed students to the LCQ and the demographic questions on the same link. The second QR code was the link to the SASNS. Students could finish the online questionnaire at their own pace within 15 minutes in the designated order. For those who finish the questionnaires quickly, they had to wait for other participants until the time was up. Lastly, the participants put away their phones and completed the receptive PV test under test conditions in 10 minutes. This long gap (approx. 20 minutes, with the VLT, the LCQ, and the SASNS completed in-between) between the productive test and the receptive test reduced the influence of testing the same PVs again consecutively (Schmitt, Dörnyei, et al., 2004). For those who finished the last test quickly, they could hand in the paper early. Participants were asked to put in their 8-digit Student ID on each instrument, which was used to identify them anonymously.

The interviews were arranged within the one-week window of the quantitative test on campus, with the IELTS speaking test taking place before the semi-structured interview. All the interviewees signed another interview informed consent form (Appendix B). The participants were encouraged to do both but they could choose to take either the interview or the speaking test. All interviews took place in a classroom, a student lounge, or the café on campus. The researcher conducted the interviews with a mobile phone to record the interview, a printed interview schedule, a pen, and a piece of paper for the interviewee to take notes in the IELTS test. The researcher emphasized that the IELTS speaking test would not be graded and it only recorded students' English production as a supplement to the standardized test. After the IELTS test, students were not given corrective feedback. When the interview was over, the researcher stopped recording and checked if the participants felt comfortable during the interview.

3.5.3 Time 2 data collection

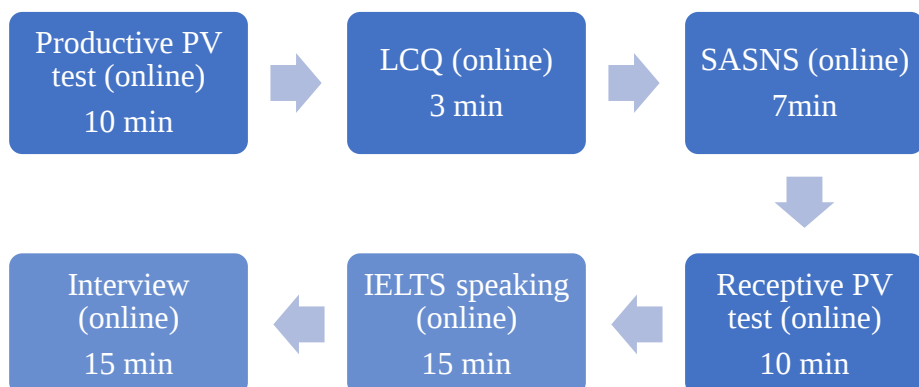
Figure 3.7b Procedure of T2 data collection



T2 data collection lasted for 40 minutes. It is 5 minutes shorter than T1 because students were exempted from filling in the demographic questions (Figure 3.7b). The researcher went to the English classes, reintroduced herself, and gave a quick recap of the study design in 5 minutes (Appendix C). Based on SASNS data in T1, the researcher listed some common misunderstandings students might have had about the definition of “English-speaking friends” and reminded the students of the available Chinese version of the questionnaires. The same PV tests were used because a three-month gap was long enough to negate *the practice effect*, while the VLT Version B was employed as an equivalent to Version A. To enhance student motivation, in each class, a 10-pound Amazon voucher was given out by a lucky draw after all the test papers were collected.

3.5.4 Time 3 data collection

Figure 3.7c Procedure of T3 data collection



At T3, data were collected remotely in two ways, due to the COVID-19 lockdown. The LCQ and the SASNS links were sent to the whole sample by email by the course leaders. The PV test links were not sent because they were unlikely to get responses based on experience from T2. Besides, the recipients might feel overwhelmed with too many tasks and the test-taking procedure could not be controlled. For the participants for whom the researcher had contact details, a time slot was arranged with each available student a few days in advance (see Figure 3.7c, 0.5 hour for non-interviewees and one hour for interviewees). To ensure test validity and reliability, the researcher supervised the test completion of each participant one-on-one by video call on WeChat or Microsoft Teams. For the few participants who preferred voice calls, the researcher asked them to set a time limit for the test and not use any dictionary. At T3, every interviewee was offered a 10-pound Amazon e-gift card to thank them for their cooperation given the extraordinary circumstances.

3.6 Data analysis

3.6.1 Research Question 1

RQ 1 probes the developmental profile of PV acquisition among international foundation learners in the UK. To answer this question, two types of data were analyzed: the PV test scores and PVs from the spoken learner corpus. Due to the low English proficiency of the participants, their spoken PV production was limited in range and number, so the test performance was deemed a more reliable parameter of their PV knowledge. As for RQ1a, the relationship between the test and corpus data and non-exposure factors (learner attributes and PV characteristics) were also investigated to reveal which variables have an influence on the PV command.

3.6.1.1 PV tests

The marked PV scores were put into Excel spreadsheets. After data import and data screening, data from demotivated participants or students cross-referencing both test

papers were discarded. The raw mean scores and mean percentage scores of PV tests, PV gains between each time point, and the VLT were calculated in SPSS. Firstly, measures of descriptive statistics were performed to check the normality of distribution. Secondly, to ascertain whether there was a significant difference between PV scores at different time points (RQ1), paired-samples *t*-tests and repeated-measures ANOVAs were performed on overall scores and PV item scores. For the non-normally distributed variables, the non-parametric equivalent analysis was also performed to cross-check the results.

A power calculator (ClinCalc, 2019) was used to determine whether the sample size at T3 has sufficient power. Results indicated that $N=28$ was needed at T3 to achieve 80% power at a significance level of $\alpha = .05$, for the *t*-test with T2 PV test data. Thus, the obtained sample size of $N = 19$ at T3 was inadequate to test the study hypothesis from the T2 sample and the results should be treated with caution for potential Type II error. It should be recognized that the sample sizes of the current study at T1 and T2 are larger than most SA empirical studies ($N=20-50$) (Borràs & Llanes, 2019; Llanes, 2011), warranting the robustness of the current results.

To answer RQ1a, non-language-contact factors were examined against PV scores: the role of semantic opacity (literal/figurative meanings) was investigated via independent-samples *t*-tests, and corpus frequency of each PV was compared against the PV item percentage scores by correlation analysis. L1 influence (Chinese/non-Chinese L1) in PV gains was investigated via MANCOVA, with the IELTS scores/VLT scores being a covariate. Students from mainland China, Taiwan, and Hong Kong were categorized as Chinese users. The role of initial and overall English proficiency in PV gains was examined via correlation analysis (against IELTS scores and the VLT scores). The decision to use VLT as a proxy of overall English proficiency was based on extensive empirical evidence of the strong relationship between them (see Section 2.2.1). Finally, the individual trajectory of PV knowledge change was examined between the three time points, and qualitative error analysis was carried out on the receptive test answers.

3.6.1.2 PV learner corpus

All IELTS speaking recordings (55 recordings, 657 minutes long, and 45330 words) were transcribed via the *O Transcribe* online software, following a simple transcription protocol based on the examples in Mackey and Gass (2015). The sentences in which the interviewees produced PVs were highlighted and extracted. Four variables were analyzed from three perspectives, being the number of PVs (types and tokens of PVs), the sophistication of PVs (operationalized as the use of low-frequency PVs), and the accuracy of PVs (operationalized as the semantically and grammatically correct use of PVs) respectively (Table 3.13).

Table 3.13 Coding protocol of PVs in the speech samples

Variables	Data type	Labels	Coded values
Number of PVs	Interval	Normalized and raw counts of types and tokens of PVs	0-1
Sophistication of PVs	Ordinal	• Most frequent 50 PVs in Liu (2011) • Most frequent 51-100 PVs in Liu (2011) • Most frequent 101-150 PVs in Liu (2011) • Not in the high-frequency PV list of Liu (2011)	• 1 • 2 • 3 • 4
Semantic accuracy of PVs	Interval	The ratio of semantically correct PVs in the speaking topic context	0-100%
Grammatical accuracy of PVs	Interval	The ratio of grammatically correct PVs in the speaking topic context	0-100%

Normalization of the PV number was calculated following the arithmetic formula in Verspoor et al. (2011), in which the largest number was counted as 1 and the smallest number as 0. As for sophistication, one of the most basic assessments of lexical sophistication of single-word vocabulary is word frequency (e.g. the frequency band of 1000, 2000, 3000, etc.) and off-list (above frequency band of 5000) words (Tracy-Ventura, 2017). Another lexical sophistication threshold of ‘rare’ words was set at the 6000 frequency level (Borràs & Llanes, 2020). In the current study, the researcher

trialed two approaches based on the high-frequency PV list by Liu (2011). The first approach differentiates frequency bands of 50, 100, 150, and off-list PVs denoting the highest PV sophistication, while the second approach only distinguishes in-list PVs (the most frequent 150) and off-list PVs. Since the results of the two approaches were similar, the result from the former approach was displayed to retain more details of sophistication. The semantic accuracy of PVs was checked with reference to Oxford Learner's Dictionary of Phrasal Verbs (McIntosh, 2006). This dictionary was chosen because it was comprehensive, up-to-date, and uses British English. *Inter-rater reliability* of PV accuracy was checked for tricky PVs with uncommon collocation (13.3% of the total produced PVs) by a native British English speaker and experienced language teacher. Non-parametric tests (Wilcoxon Signed-Rank Tests and Friedman's Test) were employed to compare data across three time points due to the non-normal distribution.

To reveal factors influencing the spoken PV production (RQ1a), the number, sophistication, semantic accuracy, and grammatical accuracy of PVs were examined against four factors: PV test scores of learners, overall English proficiency of learners, speaking topics (descriptive topics/lifestyle topics/travel topics) and task types (dialogue/monologue) of the speaking test. Correlation analysis was performed between the four PV variables and PV test scores and overall English proficiency. The role of speaking topics was examined by one-way between-groups ANOVA and Kruskal-Wallis test, while that of task types were examined via independent-samples *t*-test and Mann Whitney U test, depending on whether the dependent variables were parametric or not.

3.6.2 Research Question 2

RQ2 concerns the patterns and changes of informal L2 contact of international foundation program students in the UK and two sub-questions (2a: personal background factors that affect language contact; 2b: the influence of COVID-19 lockdown). Each sub-question was answered from two perspectives: language contact and social networks. Data from the LCQ, the SASNS, and the semi-structured interview

were analyzed to address these RQs. Since the participants had small and weak English-speaking networks, quantitative emphasis was placed on their LCQ data and qualitative analysis laid an emphasis on the reasons for their limited English-speaking friends. This finding chapter is divided into two parts, with the pre-COVID section devoted to RQ2 and RQ2a and the post-COVID section devoted to RQ2b.

3.6.2.1 Pre-COVID data

3.6.2.1a Quantitative analysis

To answer RQ2, descriptive statistics were performed on the LCQ scores at T1 and T2 and gains of the LCQ, to check the normality of distribution. Paired-samples *t*-tests were performed on overall LCQ scores and individual language contact items to ascertain whether there were significant differences between T1 and T2. The LCQ items were categorized by Exploratory Factor Analysis to reduce its dimensions. Likewise, descriptive statistics were carried out on the SASNS data. At T1 some invalid data were discarded because the friends that some respondents listed were L1-speaking or themselves. The ten variables from the SASNS data were coded based on the protocol in Table 3.14, following Dewey et al. (2013), Kennedy Terry (2017), and Xie (2017). Among them, three variables can be exported from AWS directly, whereas network density, two multiplexity measures, the frequency of contact, the intensity of friendship, and the English proficiency of friends required basic calculation. Changes in these ten parameters from T1 to T2 were compared via the Wilcoxon-Signed Rank test, due to the non-parametric nature of the variables.

Table 3.14 Coding protocol of SASNS network characteristics

	Variables	Types	Labels	Calculation	Range
1	Size (L1)	Interval	The number of friends	Exported from Part 1	0-50
2	Size (L2)	Interval	The number of friends	Exported from Part 1	0-50
3	Dispersion	Interval	The number of social groups	Exported from Part 1	0-10
4	Density	Interval	The number of connections	Calculated	0-15

			among friends and with the respondent	from Part 3	
5	Activity multiplexity	Interval	The number of social activities the learner joined in a month	The sum of the frequency in Part 4	0-100
6	Topic multiplexity	Interval	The number of topics of conversation the learner had	The sum of ticked topics from Part 5	0-15
7	Frequency of contact (online)	Ordinal	•Less than once a month •Every month •Every week •Every day	The average value of listed alters from Part 2	•1 •2 •3 •4
8	Frequency of contact (offline)	Ordinal	•Less than once a month •Every month •Every week •Every day	The average value of listed alters from Part 2	•1 •2 •3 •4
9	Intensity	Ordinal	•Not very close •Just friends •Good •Very good	The average value of listed alters from Part 2	•1 •2 •3 •4
10	English proficiency of friends	Ordinal	•Worse than me •Similar to me •Better than me •Native	The average value of listed alters from Part 2	•1 •2 •3 •4

Note: Except Variable 1, the rest are English-speaking network variables.

Data from the SASNS were also used to perform further analysis on friends' profiles, to identify the characteristics of the English-speaking friends of the sample. The following dimensions at T1 and T2 are presented descriptively, such as their means of initial encounter (in class or out of class), their nationality (same continent or different continent), their friendship intensity, and their relative L2 proficiency. Following the nationality info, a modified social network model will be put forward in the discussion.

To answer RQ2a, based on the literature review, a few participant-related potential confounders in language contact and social networks were explored: nationality, age, overall L2 proficiency, motivation, living arrangements, and center locations (Table 3.15). The influences of nationality and center locations were compared through independent-samples *t*-tests for language contact or Mann Whitney U test for social

networks, while the influence of living arrangement was compared through one-way between-groups ANOVA for language contact or Kruskal-Wallis test for social networks. The influences of overall L2 proficiency, age, and motivation were explored through Pearson correlation for language contact variables and Spearman correlation for social network variables.

Table 3.15 Analysis technique for RQ2a

Categories	Variables	Language contact	Social networks
Learner attributes	Nationality	Independent-samples t-test	Mann Whitney U test
	Overall L2 proficiency	Pearson correlation	Spearman correlation
	Motivation	Pearson correlation	Spearman correlation
	Age	Pearson correlation	Spearman correlation
External factors	Living arrangements	Between-groups ANOVA	Kruskal-Wallis test
	Center locations	Independent-samples t-test	Mann Whitney U test

It is worth mentioning that nationality was only categorized into Chinese and non-Chinese students given the idiosyncrasy of the current sample. On one hand, Chinese students were the majority in this population and a population of research interest. On the other hand, learners from other countries had very limited co-national classmates in the program, so it was presumed they would have fewer opportunities to use their L1 in social networking. Therefore, for feasibility, they were grouped as non-Chinese students for analysis. As for motivation, the score was calculated through the mean of four Ideal L2-Self items, because linguistic-related motives, rather than career-oriented motives, are more likely to increase informal L2 use during SA (Allen, 2010).

3.6.2.1b Qualitative analysis

All the interview recordings were transcribed on the *O transcribe* platform (66489 words). The Chinese interview was translated by the researcher who is a native speaker of Chinese. Thematic analysis was performed on language contact and social networks respectively on NVivo, including steps from familiarizing with transcripts, generating initial codes, searching for themes, reviewing and naming the themes, to writing up

(Braun & Clarke, 2006). Thematic analysis has the ability to summarize key features of a large amount of data and offer thick description at the same time (Braun & Clarke, 2006), but there is a long-held controversy on whether qualitative analysis can include frequency counts of codes (Elliott, 2018; Maxwell, 2010). In the current study, the main codes were counted to highlight the prominence of themes. Less salient themes were also given some attention to present the diversity of SA experiences. Two vignettes were included to illustrate cases of individual participants with different L2 social networks.

As shown in Appendix J, for language contact, the micro-level coding labeled specific L2 use activities (e.g., English reading, English emails), and personal attitudes. Then the codes were summarized into themes in the categories of overall L2 use, specific types of L2 use, and the developmental profile. In terms of social networks, the initial codes included specific types of their closest L2-speaking contacts (e.g., English teachers, classmates) and opinions toward L2 social networks. The codes were summarized into themes such as the most common way of making L2 friends and difficulties in making L2 friends. The finding chapter will present a general pattern of L2 connections, friendships with multi-nationals and host-nationals, and the developmental profile of L2 social networks. It should be noted that some Part 1 questions in the IELTS speaking test also shed light on the daily routines, hobbies, and friends of the participants, so these data could triangulate the themes summarized from the interview data.

3.6.2.2 *Post-COVID data*

RQ2b was generated against the backdrop of a pandemic in 2020, which exerted an unprecedented impact on the life of the public across the UK, including SA students. As a result of the national lockdown, the sample of this study finished the last term of study online, either self-isolating in their accommodation in the UK or studying in their home country across time differences. RQ2b was answered by data from the LCQ, the SASNS, and the semi-structured interviews at T3. Firstly, descriptive statistics of the

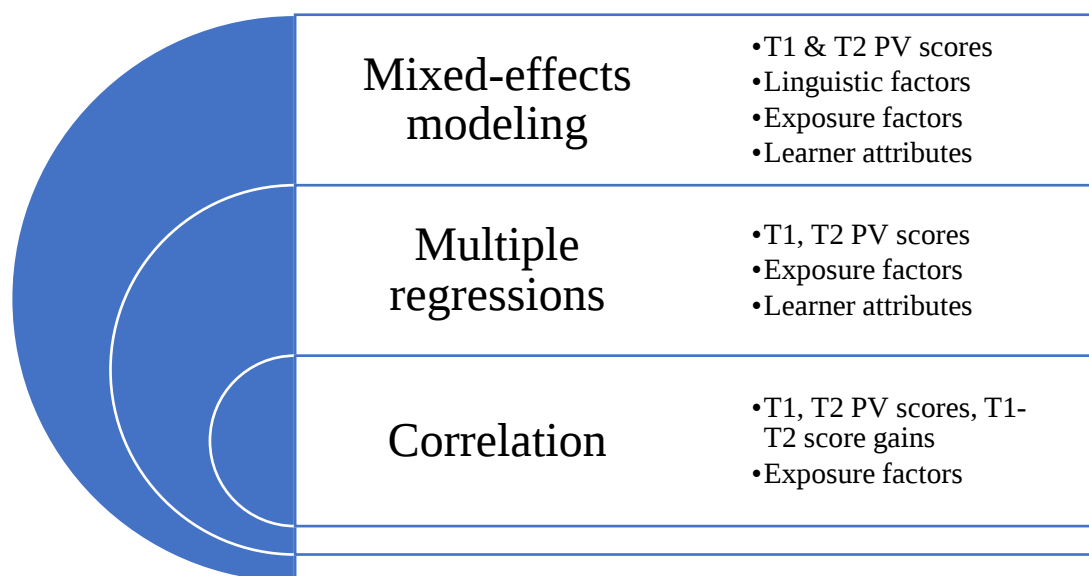
LCQ and SASNS data at T3 were provided. To compare changes across three time points, a repeated-measures ANOVA was performed on the LCQ data and a Friedman's test was conducted on the SASNS data. Lastly, thematic analysis of the interview data was presented from three aspects: language contact, social networks, and SA experience during the pandemic (Appendix J), which has a unique value in understanding the impact of COVID-19 on international students *in situ*.

3.6.3 Research Question 3

To answer RQ3 on the relationship between informal language contact and PV acquisition, inferential statistics were mainly adopted with some qualitative analyses. Firstly, three types of statistical techniques were employed to triangulate the findings from T1 and T2 data, which represent the normal SA period (Figure 3.8). Correlation analyses were performed between PV scores, PV gains, informal language contact, and social networks. Multiple regression analysis was carried out to examine the effect of informal language contact, social network, motivation, and initial/overall English proficiency (IELTS/ VLT 5k scores) on the PV scores at each time point and identify significant predictors of PV knowledge. Mixed-effects modeling was performed to examine how linguistic factors (corpus frequency, semantic opacity), exposure factors (total L2 language contact, the multiplexity of L2 social networks) and personal attributes (overall English proficiency, L1) influenced the PV scores. As a comprehensive method, mixed-effects modeling can account for most confounders that affect PV scores and it treats each answer to PV items as a case which increased the sample size considerably. Structural equation modeling is ideal for the directional relationships between the variables, but it was not employed given the inadequate sample size that completed all the instruments ($n < 100$). Ideally the data should have a minimum of 10 participants per parameter to be estimated for the equation model (Winke, 2014). Since the hypothesized model had over 20 parameters, the sample size should be larger than 200. This inadequacy makes it impractical to perform such analysis and, in hindsight, a power analysis should have been conducted beforehand to

establish a necessary sample for such analysis.

Figure 3.8 Quantitative analysis to answer RQ 3



Thematic analysis of semi-structured interviews at T3 was drawn on to elucidate how students perceived the potential benefit of language contact on PV development. Firstly, learners' attitudes toward the usefulness of PVs, the difficulty of PVs, and the ways to learn PVs were summarized, to set the scene. Secondly, regarding PV acquisition during SA, participants' opinions on the relationship between language contact and PV development were also summarized into themes, followed by a few PV acquisition examples. Lastly, some participants' reflections on the PV test scores and language contact changes were shared.

3.7 Methodological limitations

The current study entails some limitations regarding instrument design and data collection. Regarding instruments design, firstly, due to the limited duration of each data collection and the low English proficiency of participants, the number of target PVs was limited to 20, and polysemy of PVs was not accounted for. If more PVs had been tested, we could form a more comprehensive view of students' command of PVs. Secondly, to prevent *priming effects*, at the T1 and T2 interviews, students were asked

about general vocabulary and phrase learning, rather than PV-specific questions, to prevent them from cramming PVs before T2 and T3. Therefore, the amount of qualitative data on their reflections on PV learning was compromised. Thirdly, due to the diverse L1 backgrounds of students, it was not possible to prepare questionnaires in all the L1s of the participants or interview everyone in their L1. To address this issue, the researcher intentionally used simple language in the questionnaires and interviews, to avoid misunderstanding. Unlike ethnographic studies, the researcher could not live and study with the participants to carry out extensive observation of their L2 use and social circles. However, the researcher also had the advantage of being an outsider to the foundation programs and an informal ‘mentor’ of the participants, with whom they can openly express their opinions.

Regarding data collection, some advanced students from Center B did not participate in T2 and T3 data collection, because they were not required to attend English classes anymore. This number could not have been predicted before the study started, because it was unknown how many students would pass the internal English exam/meet the IELTS requirements at what stage of the year. Independent-samples t-tests were conducted to check if the dropout participants were significantly different from the remaining participants at T2. It was found that the dropout participants ($M=5.86$) had significantly higher IELTS scores than the remaining ones ($M=5.58$). It showed that the missing data could be accounted for by other observed variable (overall English proficiency). Therefore, this study took a conservative acceptance approach to deal with missing data, assuming the data were missing at random (Bhandari, 2021). The researcher was conscious not to make claims on the advanced students who only participated at T1. When performing analysis, the missing cases were deleted pairwise, to keep as many cases as possible. Future research should aim to start with a bigger sample size to account for severe participant attrition in SA research. Further, some collected PV test data at T1 had to be discarded for validity consideration, because of two different orders of distributing PV test papers.

Moreover, the interviewees recruited for the current study were predominantly Chinese

and the whole sample was selected based on *non-probability sampling*, which might limit the generalizability of the findings to SA students from other countries and other degree levels. Nevertheless, intermediate-level Chinese students make up a large percentage of the student body in British universities, so the researcher expects that the findings will still be valuable to understand the SA life of Chinese students and newly arrived international students. Besides, the administration of the LCQ might have triggered some changes to the L2-using behaviors of participants at T2 and T3. However, the researcher believes that is a positive side-effect, which can encourage students to take full advantage of their SA experience. Lastly, inter-rater reliability could be not carried out on all PV tokens in the learner corpus and the interview transcripts were not coded by a second researcher due to limited resources, which poses a threat to the validity of findings.

A series of unexpected limitations arose due to COVID-19. Firstly, the pandemic dramatically reduced the amount of T3 data and the statistical power of analysis, as the researcher could not meet the participants in person and collect data during class hours. For those who participated at T3, the overseas video call quality was sometimes unstable and the SASNS could not be accessed from mainland China. Secondly, due to the lockdown period ensuing T2 data collection, it is hard to confirm whether the lack of gains in PV knowledge at T3 happened naturally or due to the lockdown. Likewise, the patterns of L2 contact and social networks at T3 also lost their referencing value compared with normal times. If no correlation between the informal language contact variables (including social network variables) and the PV performance was observed at T3, the possible influence of lockdown measures on language development cannot be excluded. Thus, caution will be taken when drawing conclusions in the current study, especially when interpreting T3 data.

3.8 Ethical considerations

This project was reviewed by and received ethics approval through the University of

Oxford Central University Research Ethics Committee (CUREC). Evidence of CUREC approval for the study is provided in Appendix A. The present project follows the ethical guidelines of the British Educational Research Association (BERA, 2011) which outlines five primary principles of ethical research: minimizing harm; respecting autonomy; protecting privacy; offering reciprocity; and treating people equitably, with which the current study complied. The informed consent form and study information sheet can be found in Appendix B.

For instance, the time and effort of the participants were respected and acknowledged (Holliday, 2010). The arrangement of the present study aimed to cause as little inconvenience to the students as possible, as study was their priority. The quantitative data collection only took place during class hours, while the interviews were always conducted during class breaks or after class. If the students had ad-hoc meetings with their teachers or were caught up in other academic commitments, the researcher would adjust the interview slot immediately. The teachers, course leaders, and center directors who facilitated the data collection were given Amazon e-vouchers. Likewise, a 10-pound lucky draw prize and some little gifts were provided to the participants. The researcher also added the interviewees on WeChat and helped them with English-related questions or study-related questions even after the fieldwork finished. Besides, the researcher showed them around the university when they visited Oxford.

The respondents' privacy was always respected and students were told that they could withdraw from the study at any time for any reason (Dörnyei, 2007). Only student IDs were collected to match the data. All the data were anonymized and extra attention was paid to keeping their data confidential. In the design of the social network survey, considering students might not want to reveal the real names of their friends, they were allowed to write a nickname or an initial of their friend's first name.

The researcher gave high priority to students' mental well-being. For example, at T2, one interviewee showed clear signs of severe depression because she had not received an offer of a university place yet. She was suffering from insomnia and had missed two

weeks of class because of anxiety. Noticing her bad mood, the researcher gave her some consolation and tried to cheer her up instead of interviewing her as planned. At T3, this participant had received an offer and she completed all the instruments.

Moreover, an important way to acknowledge the participants is to benefit students and practitioners in some ways (Dörnyei, 2007). The researcher was invited to give a training session to 40 English teachers from 15 foundation programs across the UK on how to help Chinese students adjust to academic study at British universities. In this way, more Chinese students might benefit from better academic support from their foundation programs in the future.

CHAPTER IV FINDINGS PART I

PHRASAL VERB DEVELOPMENT

This chapter reports the findings related to RQ 1 and 1a, that is:

1. What is the developmental profile of phrasal verb (PV) knowledge among international foundation program students in the UK over one academic year?
- 1a. What non-exposure factors influence the PV scores of the participants?

Firstly, test data from over 150 students were examined through quantitative analysis, to showcase students' PV knowledge under controlled assessment. Secondly, PVs elicited from the learner corpus of 20 students were analyzed quantitatively and qualitatively in detail, offering first-hand evidence of the participants' ability to produce PVs in naturalistic discourse. The roles of some non-exposure factors in influencing PV knowledge were also explored in relation to test data and corpus data in this chapter.

4.1 TEST DATA

In this section, findings from the receptive and productive PV tests at Time 1 (T1), Time 2 (T2), and Time 3 (T3) will be reported respectively, and in comparison with each other. The descriptive statistics of test scores, score gains, relationships between test scores and some non-language-contact factors, and qualitative error analysis in the receptive PV test will each be presented in turn.

4.1.1 Descriptives and normality

This part displays the descriptive statistics of the receptive and productive PV test in a chronological order. At each time point, total PV scores, individual PV item scores, and the VLT 5k scores are presented. Normality of distribution is checked via a variety of

techniques because this information is of vital importance for future parametric analysis (Briggs Baffoe-Djan & Smith, 2019). The techniques range from the Kolmogorov-Smirnov test, the Shapiro-Wilk test, Normal Q-Q Plot, the boxplot, 5% trimmed means to eyeballing the histogram (Tabachnick & Fidell, 2013). These methods triangulated evaluations of the normality of distribution of the data, weighing the results from different measures based on Tabachnick and Fidell (2013). Overviews of all the test scores and normality checks are presented below (Table 4.1a & Table 4.1b). It should be noted that some highly proficient participants in Center B were exempted from English classes in Term 2 and a few new students enrolled in Term 2 (see Section 3.3.4 for more details), which led to some unusual score variations in this table (e.g., lower VLT score at T2 than T1). Table 4.1 below only shows the result from the full valid sample at each time point. The gains analysis on the same participants will be provided in Section 4.1.1.4.

Table 4.1a Summary of descriptives of each test from all participants at each time point

Test scores	T1			T2			T3		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Receptive PV	10.58	3.28	151	10.86	3.83	143	12.89	3.99	19
Productive PV	7.20	2.98	156	8.00	3.10	146	10.16	3.37	19
VLT 5k	15.98	7.51	160	12.89	5.89	139	/	/	/

Note: The VLT was not administered at T3.

Table 4.1b Summary of normality check of each test

	T1	T2	T3	T1-T2	T1-T3	T2-T3
Receptive PV	√	√	√	√	√	√
Productive PV	√	√	X	√	X	√
VLT 5k	√	√	/	√	/	/

Note: √ means reached normal distribution.

X means violation of normal distribution.

√ means slight violation and mostly normal distribution.

4.1.1.1 Time 1 performance

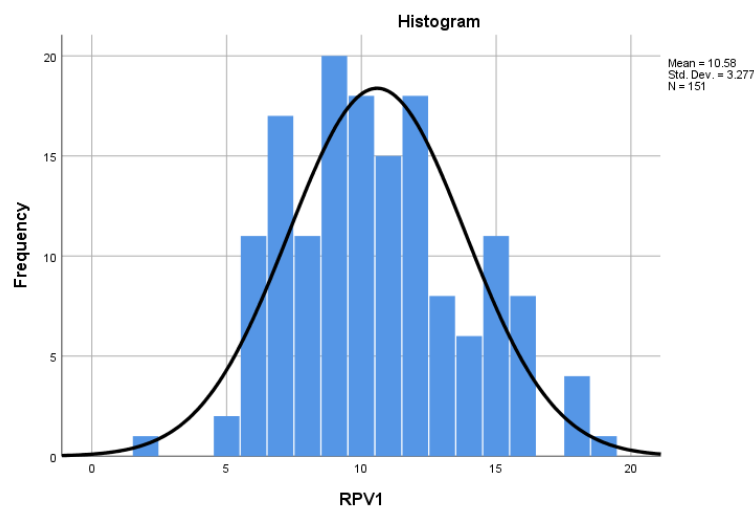
186 participants completed the T1 PV test and only the valid test papers were analyzed,

amounting to 156 cases. The possible maximum score for the receptive PV test was 20, and it is the same for the productive test. Each item was scored either 1 or 0 for correct or incorrect answers. For the VLT, the maximum score was 30 and each correct answer was scored 1.

4.1.1.1a Time 1 PV scores (raw and percentage)

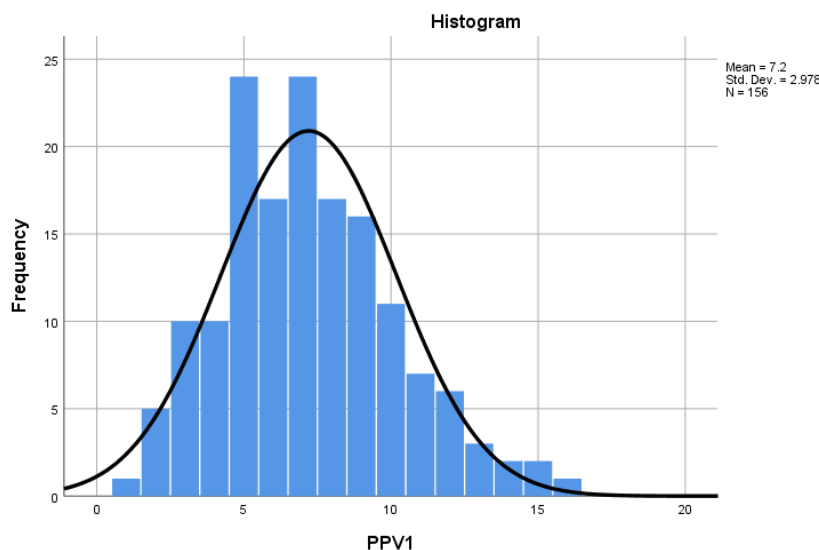
The mean of the T1 receptive test was 10.58 (N = 151, SD = 3.28), with a maximum value of 19 and a minimum value of 2. The mean percentage score was 53% of the full score. The highest percentage score was 95%, whereas the lowest percentage score was 10%. This variable was normally distributed, with skewness of .338 (SE = .197) and kurtosis of -.353 (SE = .392). The assumption of normality was checked. Firstly, the Sig. value in the Kolmogorov-Smirnov test and that of the Shapiro-Wilk test were lower than .01, suggesting a breach of normal distribution. However, according to Tabachnick and Fidell (2013), when the sample size is large, it is common to have significant results in these two tests. Therefore, they recommended eyeballing the histogram as a more appropriate criterion for normal distribution. Based on the literature, the researcher deemed this violation unproblematic, as the histogram reached good normality visually (Figure 4.1). Secondly, the Normal Q-Q Plot and the boxplot did not reveal noticeable outliers. Lastly, the 5% trimmed mean was 10.50, suggesting that cases with the top and the lowest performances did not have an impact on the mean score.

Figure 4.1: Histogram of T1 receptive PV scores



The mean score of the final dataset on the T1 productive PV test was 7.20 (N=156, SD =2.98), with the highest score being 16 and the lowest being 1. The mean percentage score was 36% of the full score, with the maximum score and the minimum score taking up 80% and 5% of the full score. This variable was normally distributed, with positive skewness of .481 (SE =.194) and a kurtosis of -.050 (SE =.386). The normality was further checked via multiple methods enumerated above, which supported the assertion of normal distribution (Figure 4.2). Inspection of the boxplot showed three outliers with the highest values among all participants. Checking these participants' profiles, the two highest-performing participants only took part in T1 data collection, which did not impact the gain analysis. The third highest-scoring participant completed data collection at three time points, which was a valuable case. Therefore, only his data were retained and the other two outliers were removed. After deletion of the two outliers, the 5% trimmed mean (7.10) indicated that trimming the highest and the lowest scores did not have an impact on the mean score.

Figure 4.2: Histogram of T1 productive PV scores



In terms of individual PV items, the mean receptive item score was .559 (SD =.40) out of a possible maximum of 1. None of the items was answered correctly by all participants. The PVs with the highest mean scores were *break down* and *come in*,

whereas the lowest-scoring items were *put out* and *pin down*. In the productive test, the mean item score was .405 (SD =.38). It can be seen that *break down* and *come in* were the top PVs that the students recalled correctly, whereas *pin down* and *make out* were answered correctly by the fewest students. The full breakdown of the descriptives of PV items is listed in Table 4.2 in the descending order of receptive scores.

Table 4.2: Descriptives of PV item scores at T1

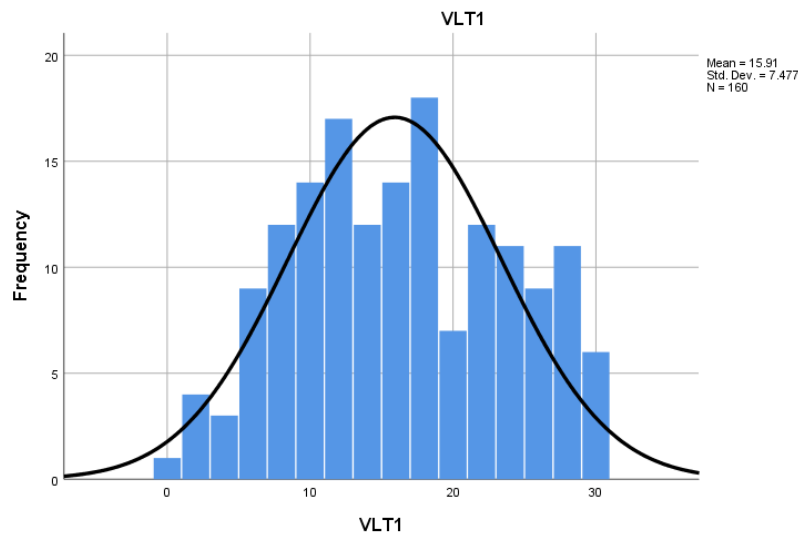
PV item	Receptive test		Productive test	
	Mean	SD	Mean	SD
Break down	0.94	0.24	0.91	0.29
Come in	0.92	0.27	0.89	0.31
Come on	0.91	0.28	0.75	0.43
Go out	0.85	0.36	0.82	0.39
Take back	0.83	0.38	0.60	0.49
Go on	0.81	0.39	0.87	0.34
Cut off	0.80	0.40	0.44	0.50
Get off	0.75	0.43	0.51	0.50
Get on	0.65	0.48	0.31	0.46
Run out	0.61	0.49	0.51	0.50
Turn up	0.47	0.50	0.33	0.47
Stand out	0.46	0.50	0.16	0.37
Take after	0.39	0.49	0.12	0.33
Make out	0.37	0.48	0.09	0.28
Hold back	0.32	0.47	0.12	0.32
Hold up	0.32	0.47	0.19	0.40
Come off	0.31	0.46	0.14	0.35
Reach out	0.26	0.44	0.17	0.38
Put out	0.14	0.35	0.15	0.35
Pin down	0.07	0.26	0.01	0.07

4.1.1.1b Time 1 VLT scores (raw and percentage)

The mean of the T1 VLT was 15.98 (SD =7.5, N =160), out of a full score of 30. The maximum score among the participants was 30, while the minimum score was 0, showing a large variance. The mean percentage score was 53.3%, with participants' performance ranging from 0% to 100%. The skewness and kurtosis values were .054 (SE=.191) and -.923 (SE=.380) respectively, indicating that the data distribution was

symmetrical but not sufficiently peaked (Figure 4.3). The histogram and the 5% trimmed mean (16.00) suggested that the data reached normal distribution, with other normality check results being considered.

Figure 4.3: Histogram of T1 VLT scores



4.1.1.2 Time 2 performance

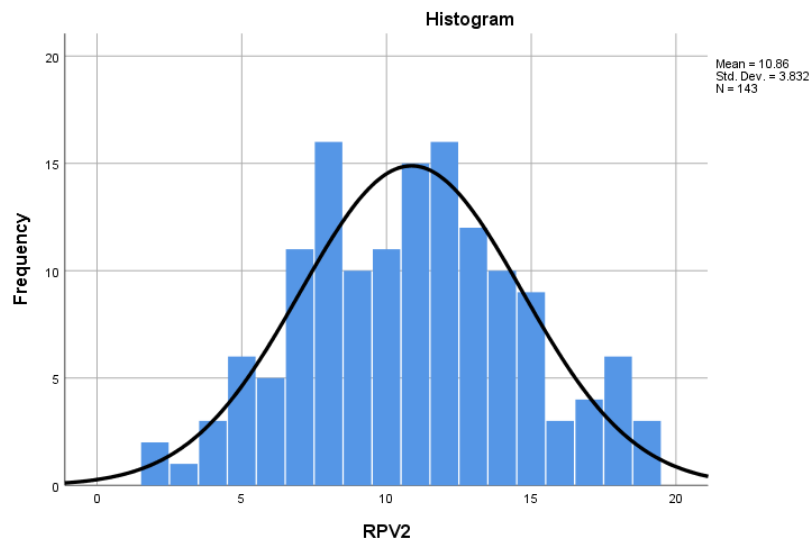
Due to participant attrition, a smaller number of participants completed the T2 data collection (N=160), and only the valid test papers (N=146) were analyzed. The same PV tests were administered, whereas Version B of the VLT 5k (see Section 3.4.2) was adopted this time.

4.1.1.2a Time 2 PV scores (raw and percentage)

The mean score of the T2 receptive PV test was 10.86 (N=143, SD=3.83), with a range of 17 (from 2 to 19). The percentage mean score was 54.3% of the full score. The highest percentage score was 95%, whereas the lowest percentage score was 10% of the total. As for distribution normality, the skewness value was .066 (SE=.203) and the kurtosis value was -.451 (SE=.403), which suggested symmetrical distribution and a relatively low peak (Figure 4.4). The Sig. value of the Shapiro-Wilk test was .116 (>.05) and the Sig. value of the Kolmogorov-Smirnov test was .026 (<.05). Since the Shapiro-

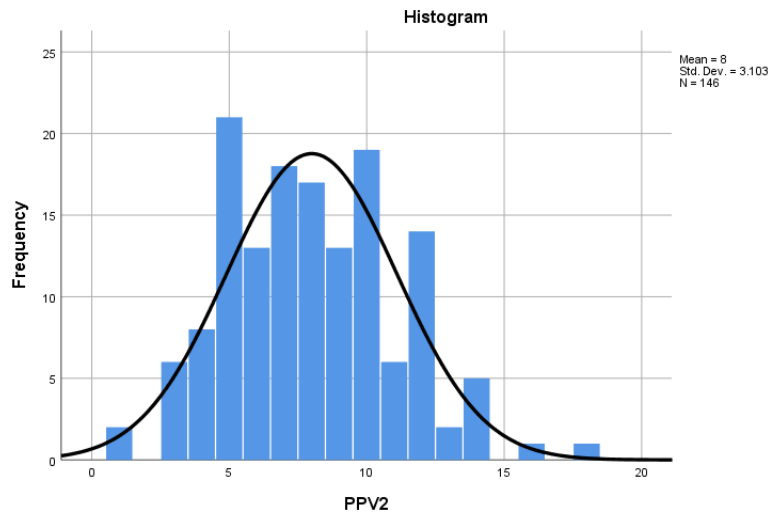
Wilk test is a stronger test for normality than the Kolmogorov-Smirnov test (Razali & Wah, 2011), this result still signals normality. This principle was also followed in the normality examination of other variables in the current study. Further inspection of the other normality check results also supported the assertion of normal distribution.

Figure 4.4: Histogram of T2 receptive PV scores



The mean score of the T2 productive test was 8 (N=146, SD=3.10), with a maximum value of 18 and a minimum value of 1. The mean percentage score was 40% of the full score. The highest percentage score was 90%, whereas the lowest percentage score was 5%. This variable was normally distributed, with skewness of .334 (SE=.201) and kurtosis of -.059 (SE=.399). The assumption of normality was checked in a few ways with conflicting results. This variable was deemed to have good visual normality (Figure 4.5), but the boxplot revealed the same noticeable outlier as the Time 1 productive test. Because he completed the test at T1 and T2, his data was retained to increase the sample size for the comparison of means.

Figure 4.5: Histogram of T2 productive PV scores



In terms of individual items, the mean score in the receptive test was .548 (SD =.42, N = 160). Most participants recalled successfully *come on* and *break down*, whereas the participants answered poorly on items such as *reach out* and *make out*. The mean item score of the productive test was .403 (SD=.36, N=160). The best-performing PVs by the participants were *come in* and *break down*. The PVs with the lowest recall rate were *pin down* and *make out*. Table 4.3 gives a breakdown of the detailed results by the descending order of the receptive PV item scores.

Table 4.3: Descriptives of PV item scores at T2

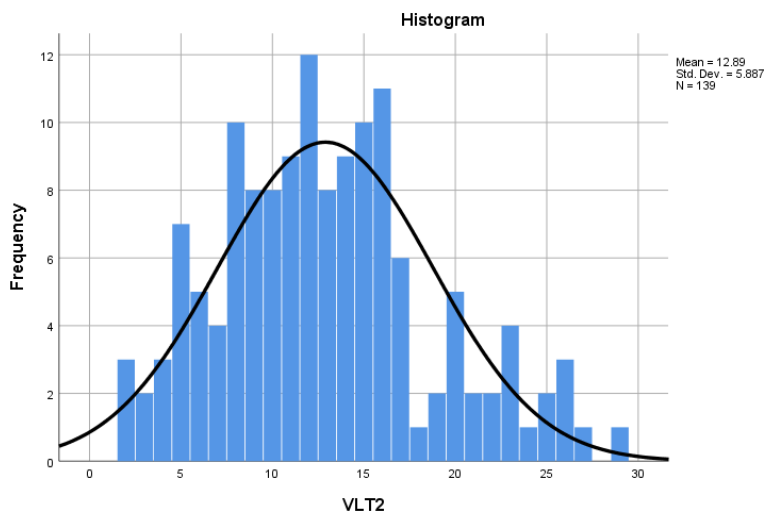
PV item	Receptive test		Productive test	
	Mean	SD	Mean	SD
Come on	0.95	0.22	0.84	0.37
Break down	0.93	0.26	0.91	0.28
Come in	0.90	0.30	0.94	0.24
Go out	0.83	0.38	0.79	0.41
Take back	0.77	0.42	0.53	0.50
Go on	0.76	0.43	0.90	0.30
Get off	0.74	0.44	0.44	0.50
Cut off	0.67	0.47	0.45	0.50
Get on	0.65	0.48	0.28	0.45
Run out	0.57	0.50	0.55	0.50
Turn up	0.51	0.50	0.38	0.49
Stand out	0.43	0.50	0.15	0.35
Hold up	0.41	0.49	0.15	0.36
Take after	0.36	0.48	0.09	0.29

Come off	0.34	0.47	0.12	0.33
Reach out	0.31	0.47	0.19	0.39
Make out	0.31	0.46	0.02	0.14
Hold back	0.26	0.44	0.11	0.32
Put out	0.16	0.36	0.22	0.42
Pin down	0.10	0.30	0	0

4.1.1.2b Time 2 VLT scores (raw and percentage)

The mean of the T2 VLT was 12.89 (N=139, SD=5.89), with a maximum recorded score of 29 and a minimum of 2. The mean percentage score was 42.97%. The minimum value was 6.67% and the maximum value was 96.67% of the total. The skewness and kurtosis values were .464 (SE=.206) and -.114 (SE=.408) respectively, indicating that the data distribution was a little skewed to the left and slightly platykurtic. The normality check suggested that data did not fully reach normal distribution. Tabachnick and Fidell (2013) did not have an ideal transformation method for the pattern in Figure 4.6. Considering the sample size was relatively large, the researcher deemed that the slight violation would not present a serious issue as regards the use of this variable in inferential analyses. The boxplot revealed two outliers that were retained because they completed T1 and T2 testing which can be used in longitudinal analysis.

Figure 4.6: Histogram of T2 VLT scores



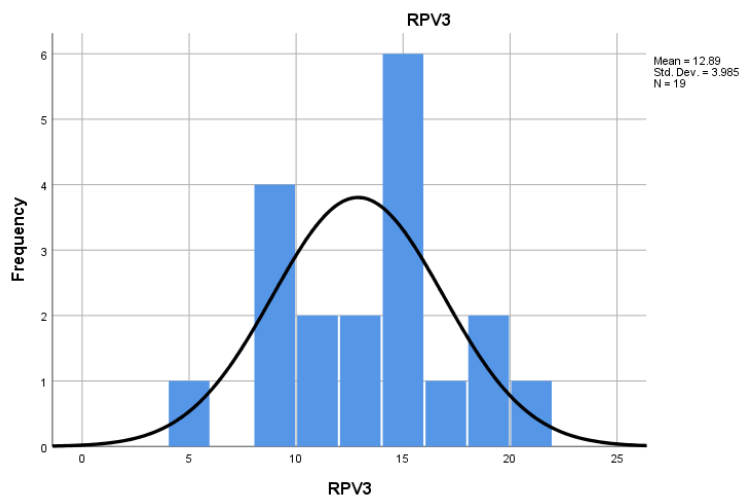
4.1.1.3 Time 3 performance

At T3, the sample size of PV tests plummeted to N=19. Based on power analysis, the T3 sample was too small to statistically represent the whole sample. Due to time constraints and issues with access to the sample, the VLT was not administered as the online format was not user-friendly and one-word vocabulary was not a primary variable of focus in the present study.

4.1.1.3a Time 3 PV scores (raw and percentage)

The mean score in the T3 receptive test was 12.89 (N=19, SD=3.99), with a maximum value of 20 and a minimum value of 5. The mean percentage score was 64.45% of the full score. The highest percentage score was 100%, whereas the lowest percentage score was 25%. This variable was normally distributed, with skewness of $-.155$ (SE=.524) and kurtosis of $-.552$ (SE=1.014). The assumption of normality was checked in various ways and the consistent results suggested a match of normal distribution.

Figure 4.7: Histogram of T3 receptive PV scores



The mean score on the T3 productive test was 10.16 (N=19, SD=3.37), with a maximum value of 18 and a minimum value of 6. The mean percentage score was 50.8% of the full score. The highest percentage score was 90%, whereas the lowest percentage score was 30%. This variable was not normally distributed, with skewness of 1.278 (SE=.524) and kurtosis of 1.433 (SE=1.014). The assumption of normality was not met (Figure

4.8a), as the boxplot and the Normal Q-Q Plot revealed two high-performing outliers. However, since they participated in three rounds of data collection, their data were retained. Based on the transformation technique in Tabachnick and Fidell (2013), this variable was log-transformed (Figure 4.8b).

Figure 4.8a: Histogram of T3 productive PV scores

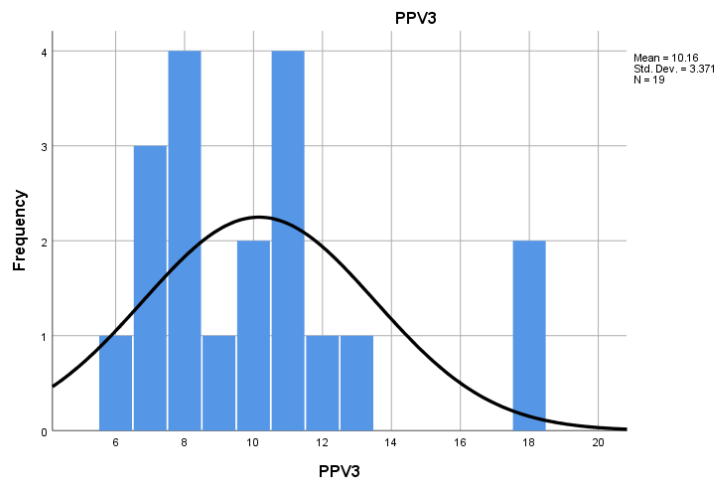
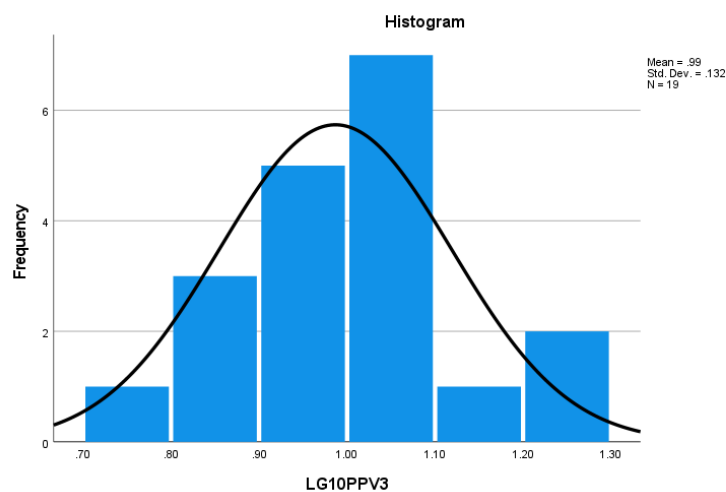


Figure 4.8b: Histogram of transformed T3 productive PV scores



When it comes to individual items, the mean receptive item score was .645 (N=19, SD=.39). All the participants answered *break down* and *come on* correctly. The remaining items with the highest means were *come in*, *go on*, and *take back*, whereas those with the lowest means were *come off*, *put out*, and *pin down*. The mean of productive item scores was .508 (N=19, SD=.34). All the participants answered Items

1 to 4 correctly, while none of the students answered item 16 correctly. The items with the highest means were *come in*, *break down*, *come on*, and *go on*, whereas the items with the lowest means were *pin down* and *make out*. The detailed descriptives of each item are provided in Table 4.4 in the descending order of the receptive scores.

Table 4.4: Descriptives of PV item scores at T3

PV item	Receptive test		Productive test	
	Mean	SD	Mean	SD
Break down	1	0	1	0
Come on	1	0	1	0
Come in	0.95	0.23	1	0
Go on	0.89	0.32	1	0
Take back	0.89	0.32	0.68	0.48
Go out	0.89	0.32	0.89	0.32
Cut off	0.79	0.42	0.58	0.51
Get off	0.74	0.45	0.53	0.51
Run out	0.74	0.45	0.68	0.48
Get on	0.74	0.45	0.42	0.51
Stand out	0.58	0.51	0.42	0.51
Make out	0.58	0.51	0.11	0.32
Turn up	0.53	0.51	0.47	0.51
Reach out	0.53	0.51	0.32	0.48
Take after	0.47	0.51	0.21	0.42
Hold back	0.42	0.51	0.26	0.45
Hold up	0.37	0.50	0.21	0.42
Come off	0.32	0.48	0.16	0.38
Put out	0.26	0.45	0.21	0.42
Pin down	0.21	0.42	0	0

4.1.1.4 Phrasal verb gains

The PV gains between each time point were calculated by the scores at the later time minus the scores at the earlier time. The gains are presented by both raw scores and percentage scores. The researcher decided to calculate and analyze gain scores because it can directly showcase the developmental profile of the participants. Since some participants were absent at either T1 or T2 data collection (see Section 3.3.4), the sample sizes for the gain scores data were much smaller. A summary of the descriptive statistics of PV gains between two time points is provided below (Table 4.5). It can be

seen that in all the pairs of comparisons there were positive gains, whose statistical significance will be shown in the next section. The large standard deviation suggested large individual variances. Further, the productive gains seemed bigger than the receptive gains, suggesting the difficulty of the receptive test. Since only T1 to T2 gains were examined on a relatively big sample, the normality checks of other gains variables (T1 to T3, T2 to T3) are presented in Appendix H. The implication of findings related to T3 will be treated with caution.

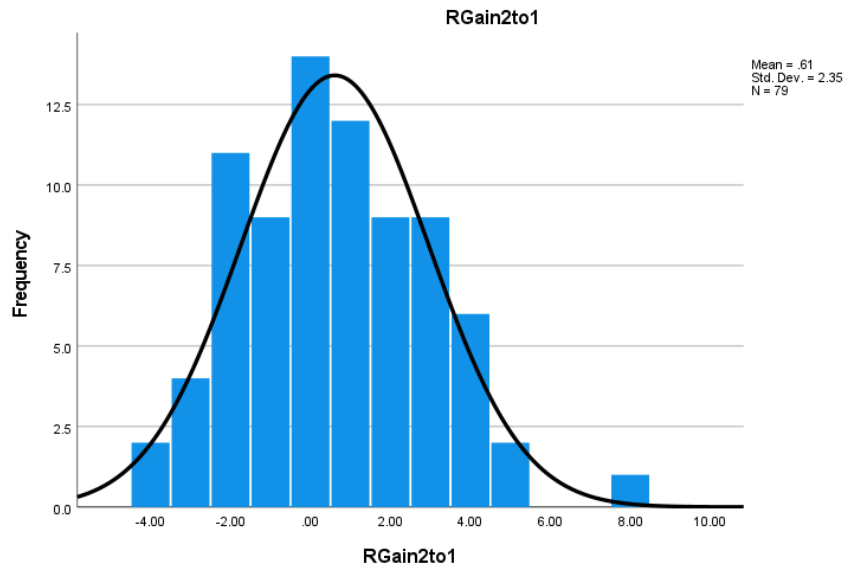
Table 4.5 Summary of PV gains between two time points

Gain	Mean	SD	N
T1-T2 receptive	.61	2.35	79
T1-T3 receptive	1.00	2.52	17
T2-T3 receptive	1.18	2.04	17
T1-T2 productive	1.35	1.96	88
T1-T3 productive	2.65	2.37	17
T2-T3 productive	.82	2.09	17

4.1.1.4a Time 1 to 2 PV gains

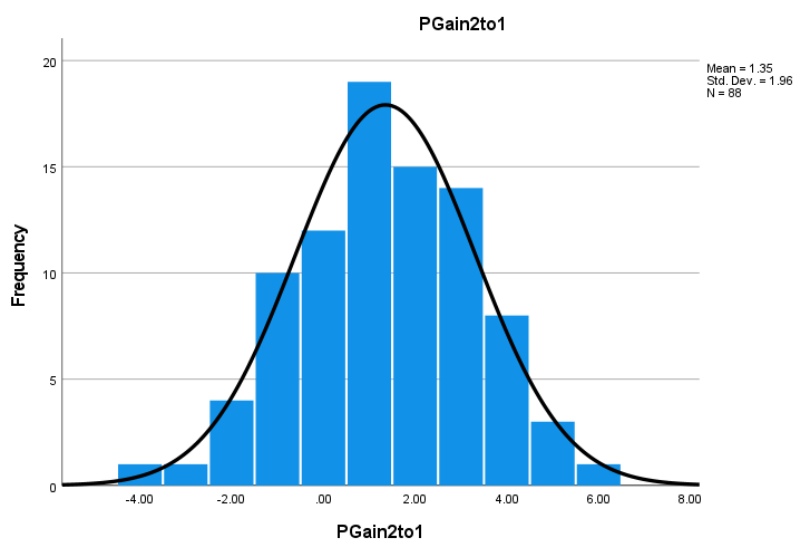
In terms of T1 to T2 receptive PV test gains, the mean was .61 (N=79, SD=2.35). The largest score decrease was -4, whereas the biggest gain was 8. The percentage gain was 3.1%, with a minimum value of -20% and a maximum value of 40%. Normality of distribution was reached based on skewness (.341, SE=.271) and kurtosis (.060, SE=.535) (Figure 4.9). Other aforementioned measures of normality also supported the assertion of normal distribution. Inspection of the Normal Q-Q Plot and the boxplot revealed one outlier with the maximum gain. Checking these data, no errors were found in the marking process or data input stage. Therefore, it was decided to retain the outlying case.

Figure 4.9 Histogram of T1 to T2 receptive PV test gains



The mean gain between T1 and T2 productive tests was 1.35 (N=88, SD=1.96), with a minimum value of -4 and a maximum value of 6, suggesting large individual differences. The percentage gain was 6.8% of the full score. The biggest score loss was -20% while the maximum gain value was 30%. The T1—T2 productive gains reached normal distribution in that the skewness value (-.134, SE=.257) and the kurtosis value (-.160, SE=.508) were close to 0 (Figure 4.10). Other means of normality check also yielded consistent results.

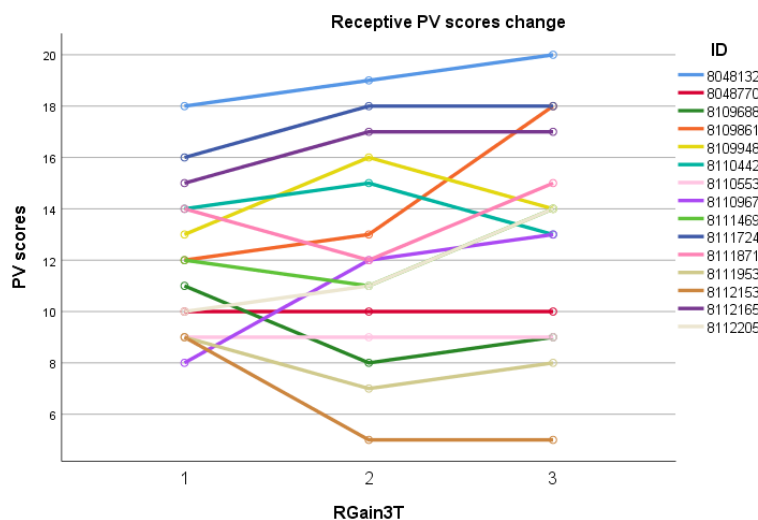
Figure 4.10 Histogram of T1 to T2 productive PV test gains



4.1.1.4b Time 1, 2, and 3 PV gains

Only 15 students participated at all three time points and the score changes of each participant are shown in the spaghetti plots below. As for the receptive tests, most participants received higher scores at T3 compared with T1, with four cases achieving scores lower than T1 (Figure 4.11). Nine participants made steady changes (only one-way change or plateau), whereas six participants experienced fluctuation in the scores to some extent (i.e., downward and upward trend). Among this sample, the mean T1-T2 receptive gain was .20 (SD=2.24), suggesting huge individual variance. The maximum gain was 4 and the biggest score loss was -4. The mean T2-T3 receptive gain was .93 (SD=1.91). The maximum gain was 5 and the maximum loss was -2. T1-T2 gains and T2-T3 gains both reached normal distribution in this sample.

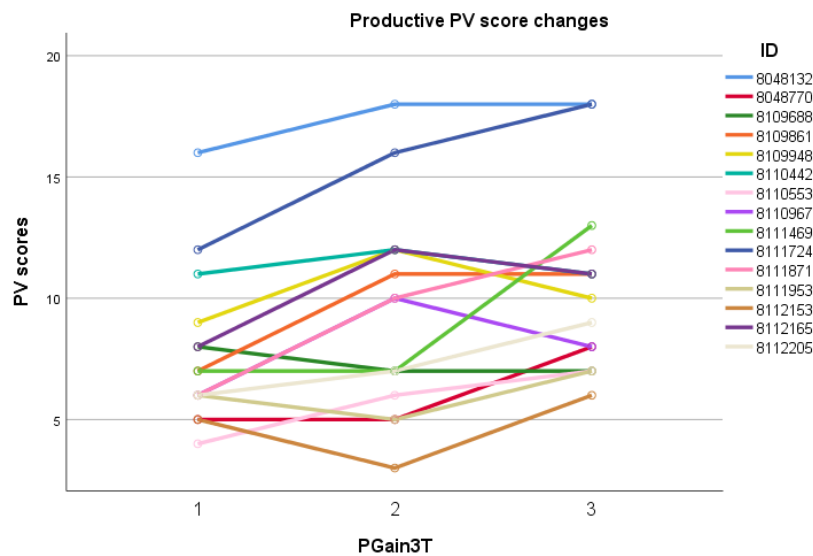
Figure 4.11 Spaghetti plot of Time 1, 2, and 3 receptive test scores



In the spaghetti plot of the productive test, all participants achieved higher scores at T3 than at T1, except one case who scored the same at T3 as at T1 (Figure 4.12). Eight participants made steady gains (only upward trend or plateau) across all time points, whereas seven participants experienced fluctuation in the scores to some extent (i.e., having both downward and upward trends between all time points). Among this subsample, the mean T1-T2 productive gain was 1.67 (SD=2.13). The maximum gain was 4 and the maximum loss was -2. The mean T2-T3 productive gain was 1 (SD=2.17). the maximum gain was 6 and the minimum gain was -2. T1-T2 gains and T2-T3 gains

both reached normal distribution in this sample.

Figure 4.12 Spaghetti plot of Time 1, 2, and 3 productive test scores



4.1.1.5 Summary of descriptives and normality

The descriptive statistics and normality check of the receptive and productive PV test scores and the VLT scores indicate that these tests were adequate to distinguish students of different levels in the sample, though many participants did not obtain high marks, even at T3. The tests were effective, and the *ceiling effect* of repeated testing was avoided. Receptive PV scores were higher than productive PV scores throughout. In the examination of normality, although sometimes the Sig. values of the normality tests were significant, most histograms reached normal or almost-normal distribution by other evaluations such as eyeballing. The only variable that was transformed was the T3 productive PV scores, the transformation of which resulted in a normal distribution. To account for the non-normal distributions of this variable, when a parametric test was carried out on it, the equivalent non-parametric test was also conducted and was reported in order to demonstrate the convergence or divergence of the findings of the two analytical approaches.

4.1.2 Comparison of means

Paired-samples *t*-tests and repeated-measures analyses of variance (ANOVAs) were used to compare the difference in the PV performances of learners. Due to COVID-19, the sample size at T3 was extremely small. To increase the sample size of the tests, *t*-tests were used for comparisons between T1 and T2, instead of using the post-hoc tests in ANOVAs for all comparisons. Bonferroni correction was adopted in the following analysis to reduce the chance of Type I error. For each dataset, the standard *p*-value threshold (.05) was divided by the number of analyses that were conducted on that dataset. However, Bonferroni correction is often criticized for being too conservative and it increases the chance of Type II error (Nakagawa, 2004). Therefore, the original significance is also shown in the chapter or Appendix H for reference.

4.1.2.1 Performance between types of PV knowledge

Independent-samples *t*-tests were conducted on the mean scores of the receptive and productive PV tests at the same time point, to demonstrate whether productive knowledge preceded receptive knowledge in this sample. It also partly answers RQ1, because it mapped the developmental trajectory of PV acquisition by comparing the gap between learners' receptive and productive knowledge. The mean difference was calculated by receptive scores minus productive scores. Bonferroni correction was applied, so the difference was considered significant only when the *p*-value was smaller than .0025 (.05/20). Non-significant results implied that learners knew the PV equally well in both the productive and the receptive forms or did not know the PV in either test. Overall, the item differences at T1 and T2 were consistent and receptive scores were consistently higher than productive scores with very few exceptions (Table 4.6). At T3 due to the small sample size, there was a less significant difference.

At T1, there were statistically significant differences between the productive and receptive scores of 10 PVs. For the remaining items, the participants had good productive and receptive command of frequent PV items, such as *come in*, *go on*, *go*

out and *break down* ($p > .05$), whereas they did not master *put out* in either the productive or the receptive test ($p > .05$). At T2, statistically significant differences were observed in 12 PVs. At T3, the significant difference was only observed in one item, which is *make out*. Comparing results between three time points, the biggest mean differences were seen on less frequent PVs, such as *get on*, *make out*, *take after*, and *stand out*, while other highly frequent PVs were recalled accurately both receptively and productively.

Table 4.6 Comparison of means between productive and receptive PV tests

PV	T1 Mean Difference	Significance (p)	T2 Mean difference	Significance (p)	T3 Mean difference	Significance (p)
Cut off	0.36	0.000**	0.22	0.000**	0.21	0.172
Get on	0.34	0.000**	0.37	0.000**	0.32	0.05
Stand out	0.31	0.000**	0.29	0.000**	0.16	0.344
Make out	0.28	0.000**	0.29	0.000**	0.47	0.002**
Take after	0.27	0.000**	0.27	0.000**	0.26	0.092
Get off	0.24	0.000**	0.3	0.000**	0.21	0.188
Take back	0.23	0.000**	0.23	0.000**	0.21	0.119
Hold back	0.2	0.000**	0.15	0.001**	0.16	0.318
Come off	0.17	0.000**	0.22	0.000**	0.16	0.265
Come on	0.16	0.000**	0.1	0.003	0	N/A
Turn up	0.14	0.007	0.13	0.02	0.06	0.754
Hold up	0.12	0.011	0.26	0.000**	0.16	0.296
Run out	0.1	0.05	0.02	0.773	0.06	0.729
Reach out	0.1	0.036	0.12	0.015	0.21	0.199
Pin down	0.07	0.003	0.1	0.000**	0.21	0.042
Come in	0.03	0.302	-0.04	0.221	-0.05	0.331
Break down	0.03	0.321	0.01	0.648	0	N/A
Go out	0.03	0.479	0.04	0.376	0	1
Put out	0.003	0.933	-0.07	0.132	0.05	0.712
Go on	-0.06	0.135	-0.14	0.001**	-0.11	0.163

Note: **Statistically significant after Bonferroni correction

Significance N/A when productive score = receptive score

4.1.2.2 Vocabulary knowledge scores at Time 1 and Time 2

Paired-samples *t*-tests were conducted on the total PV test scores to evaluate the changes in the participants' performance between T1 and T2. Firstly, there was a

statistically significant gain in receptive scores from T1 (M=10.32, SD=3.13) to T2 (M=10.92, SD=3.77): $t(78) = 2.298, p=.024, N=79$. The mean receptive gain was .608, with a 95% confidence interval ranging from .081 to 1.134. Partial η^2 is one of the most common effect size statistics to compare groups, which indicates the proportion of variance of the dependent variable that is explained by the independent variable (Pallant, 2020). The partial η^2 (0.06) indicated a moderate effect size. Second, there was a statistically significant gain in productive PV scores from T1 (M=6.61, SD=2.69) to T2 (M=7.95, SD=3.06): $t(87) = 6.403, p<.001, N=88$. The mean productive gain was 1.341, with a 95% confidence interval ranging from .925 to 1.757. The partial η^2 statistic (0.31) indicated a large effect size. Lastly, there was no significant change in the VLT scores from T1 (M=15.38, SD=6.98) to T2 (M=14.28, SD=5.86): $t(80) = 1.912, p=.059, N=81$. To sum up, the sample improved their PV knowledge significantly between T1 and T2 in terms of receptive scores and productive scores. There was no significant difference in lower-frequency single-word vocabulary knowledge as measured by the 5k level of the VLT, between T1 and T2.

4.1.2.3 Phrasal verb knowledge scores at Time 1, 2, and 3

In this section, one-way repeated-measures ANOVAs were performed on receptive PV scores and productive PV scores across all time points respectively. The sample size of these analyses was 15, due to the pandemic impact at T3. Although this small sample has limited statistical power, ANOVA was still performed to track the longitudinal changes of these participants over seven months.

Between T1 (8 weeks after arrival), T2 (20 weeks after arrival), and T3 (30 weeks after arrival), in the receptive test, no statistically significant gain was observed in the repeated-measures ANOVA (Wilk's Lambda=.772, $p=.186, N=15$). The descriptives at each time points were: T1 (M=12.00, SD=2.95), T2 (M=12.20, SD=4.14), T3 (M=13.13, SD=4.24). In the repeated-measures ANOVA for the productive test, there was a significant effect for time, Wilks' Lambda=.373, $[F(1, 14) = 10.916, p=.002, N=15]$, multivariate partial η^2 being .627, indicating a large effect size. Post-hoc

comparisons indicated that there were two pairs of significant differences. From T1 (M=7.73, SD=3.17) to T2 (M=9.40, SD=4.22), the mean gain was 1.67, $p=.027$. From T1 (M=7.73, SD=3.17) to T3 (M=10.40, SD=3.72), the mean gain was 2.67, $p=.001$. T2 productive scores did not have a significant difference from T3 productive scores ($p=.288$). Since T3 productive score was not normally distributed, Friedman's test was also conducted. The result was in line with the repeated-measures ANOVA that there was a main effect of time: $\chi^2(2) = 11.593, p=.003$.

4.1.2.4 Summary of comparison of means

A number of analytic techniques were employed in this section to probe mean differences cross-sectionally and longitudinally. Independent samples *t*-tests revealed that the majority of tested PVs yielded significantly different scores in productive and receptive tests at T1 and T2, suggesting that students had different degrees of receptive and productive knowledge. Paired-samples *t*-tests, one-way repeated-measures ANOVAs, and Friedman's tests were employed to examine changes of the means in the receptive and productive PV scores. After Bonferroni correction, between T1 and T2, productive PV scores and receptive PV scores both saw significant gains, and these gains could not be traced back to specific PV item gain. On the sample that completed three rounds of data collection, the significant gain was only observed in productive scores. It should be noted that between T1 and T2 the sample size was larger and was not impacted by the pandemic. The significant gains during that period might be a more generalizable result for the SA learners studying under normal circumstances. A recapitulation of the comparison of means results on the test scores between time points is listed below (Table 4.7).

Table 4.7 Tests with significant gains over time points after Bonferroni correction

Test scores	T1-T2	T1-T3	T2-T3	T1-2-3
Receptive	√			
Productive	√	√		√

4.1.3 Non-exposure factors that influence the phrasal verb knowledge

In this section, the roles of non-language-contact factors on the PV scores were explored from the PV perspective (e.g., semantic opacity, corpus frequency) and the learner perspective (e.g., the L1 and the overall English proficiency of the participants). These four factors for investigation were selected based on existing literature. PV is a complicated linguistic structure with varied opacity, frequency, transitivity, and separability (Kamarudin et al., 2019), and PVs were harder for lower-level learners and learners whose L1 did not possess similar structures (Liao & Fukuya, 2004). The present study took a step forward by not only looking at their influences on PV knowledge (raw test scores) but also PV gains (test gain scores). Examining these factors separately also contributes to answering RQ 3 (the relationship between PV and language contact). Through these initial analyses, we can know what factors might be covariates and what factors should be put in the regression models and mixed-effects models. The analysis techniques and results summary for each factor are listed below (Table 4.8).

Table 4.8 Summary of non-language-contact factors analysis techniques

Factor	Analysis techniques	Significance
Semantic opacity	<i>T</i> -test	No
Corpus frequency	Correlation, line graph	Yes
Overall English proficiency	Correlation	Yes
L1	<i>T</i> -test, one-way MANCOVA	No

4.1.3.1 Semantic opacity

Independent-samples *t*-tests were performed on PVs of figurative meaning and literal meanings (see Table 3.7 for the list of PV meanings) in both tests at each time point. Although literal meanings had higher average scores than figurative meanings at a descriptive level, *t*-tests indicated that there was no significant difference in either the receptive or productive test at T1, T2, or T3 ($p > .05$) (Appendix H). Neither was any significance observed in the receptive or productive gains of the participants who

participated three times. Since T1-T3 productive gain did not reach normal distribution, a non-parametric equivalent Mann-Whitney U test was conducted. The same non-significant result was yielded. Therefore, this factor was not identified as a key factor influencing the PV command of the selected PVs.

4.1.3.2 Corpus frequency

To determine whether a relationship existed between PV corpus frequency, PV scores, and PV gains, correlation analysis was performed following the practice in Schmitt and Redwood (2011). A premise of using the parametric correlation technique is the normal distribution of variables. The normal distribution of PV scores was reported in Section 4.1.1. Non-parametric analysis was adopted because the BNC and COCA corpus frequency in Liu (2011) of the selected PVs yielded significant Sig. values ($p < .001$) in the Kolmogorov-Smirnov test and the Shapiro-Wilk and the histogram suggested skewed distribution toward high-frequency PVs. Spearman's correlation was carried out and the results are presented in Table 4.9. Mean PV scores of all participants in each test and mean PV scores of the participants who completed the tests three times were both used in the analysis and the correlation coefficients were similar, so results on the latter subsample are presented to avoid repetition.

Overall, the corpus frequencies of PVs had a significant strong relationship with their mean scores, with three main findings. Firstly, the BNC frequency has a stronger correlation than the COCA frequency, which might be attributed to the fact that students were living in the UK. Secondly, although both productive scores and receptive scores were significantly correlated with corpus frequency, the effect size was bigger in productive knowledge, indicating that productive knowledge was more directly associated with frequency. Lastly, the receptive and productive PV gains did not yield any significant correlation with the corpus frequency, showing that this factor did not influence whether the knowledge of target PVs was improved or not.

Table 4.9 Spearman correlation between corpus frequency and PV scores

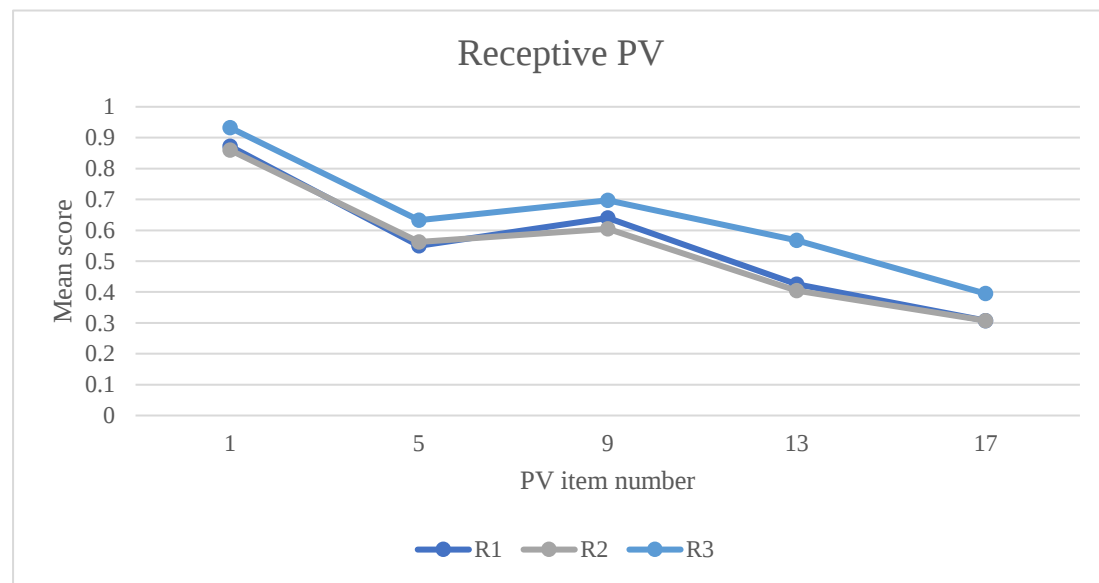
N=18	COCA frequency		BNC frequency	
PV Scores	ρ	p	ρ	p
T1 productive	.671**	.001	.718**	.001
T1 receptive	.471	.049	.647**	.004
T2 productive	.668**	.002	.743**	.000
T2 receptive	.516	.028	.649**	.004
T3 productive	.592**	.010	.677**	.002
T3 receptive	.468	.050	.607**	.008
PGainT2-T1	.263	.291	.106	.675
PGainT3-T1	.135	.593	.076	.763
PGainT3-T2	.366	.136	.309	.213
RGainT2-T1	.276	.268	.097	.702
RGainT3-T1	.105	.677	.167	.507
RGainT3-T2	.159	.529	.187	.458

Note: P=productive, R=receptive

**Significant at $p < .01$ level

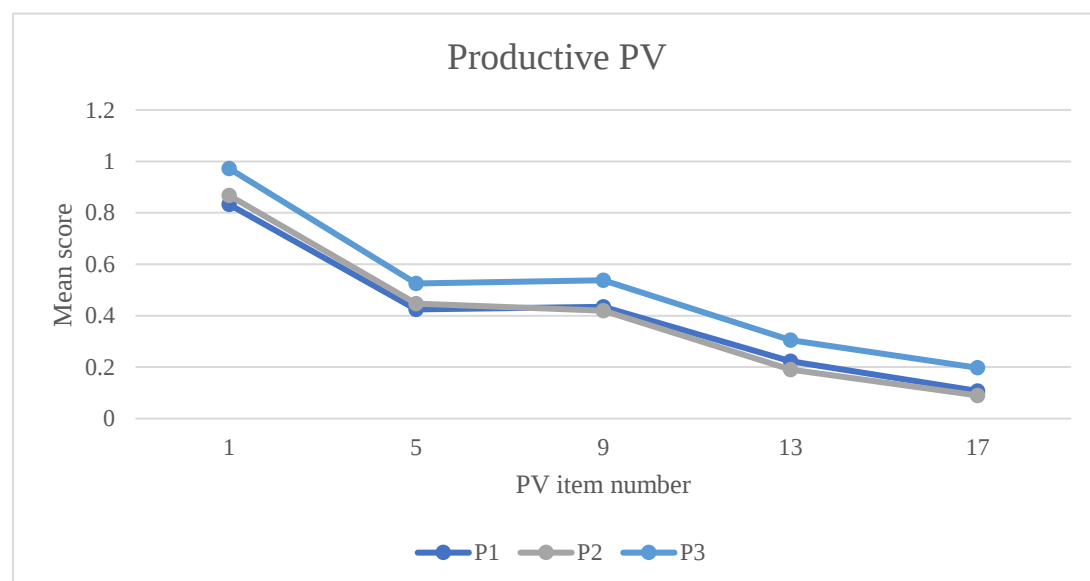
Another way to examine the role of PV frequency is visual presentation, which can triangulate the correlation findings. The 20 target PVs' mean percentage scores were sorted into groups of four (to even out extreme item variation) by their corpus frequency in the BNC and COCA. The moving averages of each group at each time point were calculated and shown. Two sets of data were used: data from all the participants at each time point and data from those who took the tests three times. The line graph patterns were basically identical between the smaller sample and the larger sample, showing the reliability of the relationship. To avoid repetition, only the BNC-referred line graphs from the participants who completed the test three times are presented below (Figures 4.13a and 4.13b). In both tests, as corpus frequency declines, the mean scores decrease accordingly. For example, the mean receptive score of the 1st to the 4th most frequent PVs in the BNC was higher than the mean scores of the 5th to the 8th most frequent PVs in the BNC. Overall, the participants mastered the top 20 most frequent PVs in the BNC, whereas their command of the top 30-100 most frequent PVs was similar. Their knowledge of the less frequent PVs was the weakest.

Figure 4.13a Receptive PV mean scores trend by descending corpus frequency order



Note: R= receptive test

Figure 4.13b Productive PV mean scores trend by descending corpus frequency order



Note: P= productive test

4.1.3.3 English proficiency

Research has shown that general vocabulary knowledge is usually proportional to English proficiency (Schmitt, 2014). Therefore, correlation analysis used initial IELTS scores (collected at T1), and VLT 5k scores at T1 and T2 as proxies of English proficiency, to examine their relationships with PV knowledge. It should be noted that

some students did not participate in both T1 and T2 data collection. Thus, when performing cross-time-point correlation, the sample sizes were smaller. PV knowledge variables include receptive and productive PV scores at three time points and gain scores between three time points. Especially, extra analyses were performed on two PV variables without normal distribution. Log-transformed T3 productive score was also correlated with the independent variables, showing similar correlation coefficients as the original variable. T1-T3 productive PV gain was trialed with Spearman correlation, the result suggested a positive correlation with VLT scores at T2 ($\rho = .612$, $p = .015$), consistent with the Pearson correlation result.

Pearson correlation demonstrated a positive relationship between English proficiency and PV scores (Table 4.10). After Bonferroni correction (p threshold $= .05/15 = .0042$), T1, T2, and T3 PV scores still showed significant correlations with English proficiency. In terms of IELTS scores, there were moderate correlations with T1 and T2 productive and receptive scores. The strongest correlation was observed between IELTS and T3 productive and receptive scores. There was no significant correlation between IELTS scores and PV gain scores. English proficiency (measured by VLT2) might have a marginal positive relationship with the development rate of productive PV knowledge, although the significance was just above the threshold (T1-T2 productive gain: $r = .310$, $p = .005$).

Table 4.10 Pearson correlations between overall English proficiency and PV scores

Variables	IELTS score			VLT1			VLT2		
	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N	<i>r</i>	<i>p</i>	N
T1 productive	.360**	.000	92	.459**	.000	124	.452**	.000	84
T2 productive	.469**	.000	74	.486**	.000	82	.573**	.000	150
T3 productive	.816**	.004	15	.531	.023	18	.703**	.002	17
T1 receptive	.461**	.000	86	.524**	.000	121	.442**	.000	78
T2 receptive	.493**	.000	72	.539**	.000	83	.608**	.000	145
T3 receptive	.690	.010	15	.651**	.003	18	.641	.006	17
Pgain T1-T2	.272	.049	53	.177	.161	64	.310	.005	81
Pgain T1-T3	.447	.109	14	.081	.757	17	.551	.033	15
Pgain T2-T3	.172	.575	13	-.258	.335	16	.162	.535	17

Rgain T1-T2	.066	.651	49	.238	.065	61	.242	.040	73
Rgain T1-T3	.290	.315	14	.319	.212	17	.279	.314	15
Rgain T2-T3	.340	.256	13	.046	.864	16	.160	.541	17

Note: P=productive, R= receptive

**Statistically significant after Bonferroni correction

4.1.3.4 L1 influence

Another factor that might influence the developmental profile of PV knowledge is L2 learners' L1. Therefore, independent-samples *t*-tests were performed on T1 and T2 data, to confirm whether Chinese and non-Chinese students had statistically significant differences in their PV scores. At T3, the number of non-Chinese students was only two, which was too small for meaningful comparison. This dichotomous categorization was made because many countries only had one participant in the sample. There were over 110 Chinese participants and over 30 non-Chinese participants in T1 and T2 data collections (see Table 3.3 for a nationality breakdown). At a descriptive level, non-Chinese students had higher mean PV scores in general.

After Bonferroni correction, only those with $p < .0056$ (.05/9) were considered statistically significant. Significant advantages were observed on non-Chinese participants in the VLT scores [$t(149) = -5.632, p < .001$] at T1, receptive scores [$t(57.404) = -4.267, p < .001$] and VLT scores [$t(141) = -3.188, p = .002$] at T2. Regarding gain scores, non-Chinese students had non-significant advantages in T1-T2 receptive and productive PV gains. Interestingly, Chinese students only outperformed non-Chinese students in terms of T1-T2 VLT gains: [$t(79) = 3.049, p = .003$], suggesting they paid more attention to one-word vocabulary. Overall, non-Chinese students had larger general vocabulary size and higher receptive PV scores at T2, while Chinese improved general vocabulary size faster than non-Chinese students. The details of these comparisons are listed below (Table 4.11).

Table 4.11: Independent-samples *t*-tests in vocabulary scores between learner groups

Variable	L1	N	Mean	SD	t	p
T1 receptive	Chinese	108	10.15	3.29	-2.093	.038
	non- Chinese	36	11.47	3.28		

T1 productive	Chinese	113	6.97	2.85	1.120	.265
	non- Chinese	36	7.61	3.34		
T1 VLT	Chinese	98	13.52	6.69	-5.632	.000**
	non- Chinese	53	20.06	7.03		
T2 receptive	Chinese	115	10.43	3.94	-4.267	.000**
	non- Chinese	30	13.27	3.03		
T2 productive	Chinese	118	7.73	3.16	-2.739	.007*
	non- Chinese	32	9.41	2.72		
T2 VLT	Chinese	111	12.37	6.03	-3.188	.002**
	non- Chinese	32	16.19	5.76		
T1-T2 receptive gain	Chinese	63	.32	2.37	-2.163	.034
	non- Chinese	14	1.79	1.93		
T1-T2 productive gain	Chinese	72	1.14	1.83	-2.215	.029
	non- Chinese	16	2.31	2.30		
T1-T2 VLT gain	Chinese	53	.12	3.34	3.049	.003**
	non- Chinese	24	-3.46	3.54		

Note: ** Statistical significance after Bonferroni correction

*Approaching statistical significance after Bonferroni correction

However, it is possible that the significant between-group differences were attributed to their different English proficiency in the first place. Therefore, one-way MANCOVAs were carried out on T1 and T2 data respectively. L1 was taken as the independent variable, with productive PV scores and receptive PV scores as two separate dependent variables and IELTS scores/VLT scores as a covariate. Since only two participants at T3 were not Chinese, this analysis was not performed on T3 data.

A number of assumptions were checked prior to performing the analysis. The sample had the independence of observations, and the sample size was bigger than the number of dependent variables. Scatterplot matrices confirmed the linearity between the covariate and the dependent variables. There was homogeneity of regression slopes of PV scores by checking the grouped scatterplot. Homogeneity of variance-covariance matrices was met by checking the Sig. value in the Box's M Test ($p > .05$). Additionally, univariate normality and multivariate normality were reached because there was no significant outlier in either the receptive or the productive PV test scores, based on the Mahalanobis Distance value in the regression models. Other single-variable normality

checks performed in Section 4.1.1 also proved the normality. Lastly, multicollinearity and singularity were absent by checking the correlation coefficients of the two dependent variables.

The results were consistent at T1 and T2, no matter whether the analysis used VLT or IELTS scores as the covariate. There was no statistically significant difference between the Chinese and non-Chinese participants on the combined dependent variables of PV command at T1 and T2 after controlling for their IELTS/VLT scores. At T1, $F(2, 108) = 1.299$, $p = .277$, Wilks' $\Lambda = .977$, partial $\eta^2 = .019$. Similarly, at T2, $F(2, 131) = 1.702$, $p = .186$, Wilks' $\Lambda = .975$, partial $\eta^2 = .023$. Although there was no need to conduct one-way ANCOVA based on this non-significant result (Tabachnick & Fidell, 2013), the researcher nevertheless conducted four one-way ANCOVAs on productive PV scores and receptive PV scores at T1 and T2 separately and no significant influence of L1 was detected. To sum up, the one-way MANCOVA showed the decisive role of overall English proficiency in predicting PV scores. Thus, any L1 disadvantage to Chinese students could not be established from the current sample.

4.1.3.5 Summary of non-language-contact factors affecting PV scores

This section intends to find out what non-exposure factors might affect PV scores and PV gains, to lay a foundation for the more complex model building in RQ3. Based on the results above, corpus frequency and overall English proficiency showed strong relationships with the receptive and productive PV scores: higher corpus frequency of a PV and higher English proficiency of the participant were associated with better PV knowledge at different time points. Whether the participants had a Chinese L1 did not influence PV knowledge directly. Likewise, the semantic opacity of PVs did not affect their acquisition significantly, suggesting that the influence of corpus frequency overrides the opacity of figurative meanings. As the gain on each PV item was very small (usually $<.02$), semantic opacity and corpus frequency of the target PVs did not show significant influence on individual PV item gains. Close to the significance threshold, more proficient students might have a little advantage in productive PV gains.

4.1.4 Further analysis of phrasal verb tests

The present study yielded a large amount of data on students' command of PVs at different time points. Based on data from the sample that participated in all data collection points (n=15), two analyses were carried out further.

4.1.4.1 Changes in the degree of knowledge

Firstly, students' knowledge of each PV in two PV tests was coded at three levels: unknown, partial knowledge, and full knowledge (see Figure 4.14) and was compared cross-sectionally (between PVs). Then, their longitudinal changes were coded at four levels: gains, stabilization, attrition, and fluctuation (see Figure 4.15), to directly addresses RQ1 of the developmental profile of PV knowledge over the seven-month data collection period.

Figure 4.14 Coding of the PV knowledge level

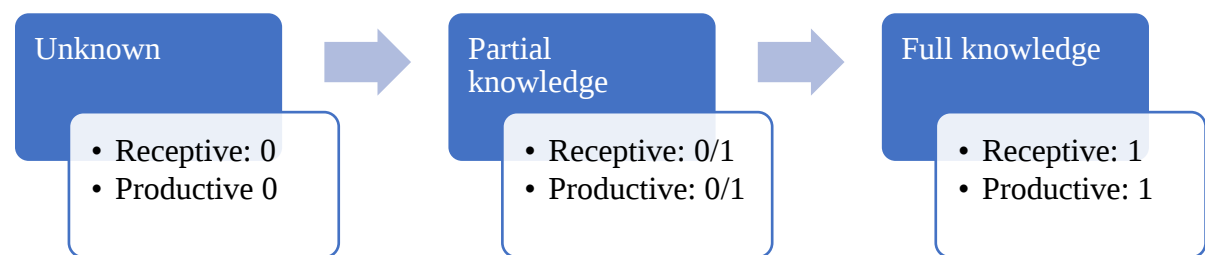


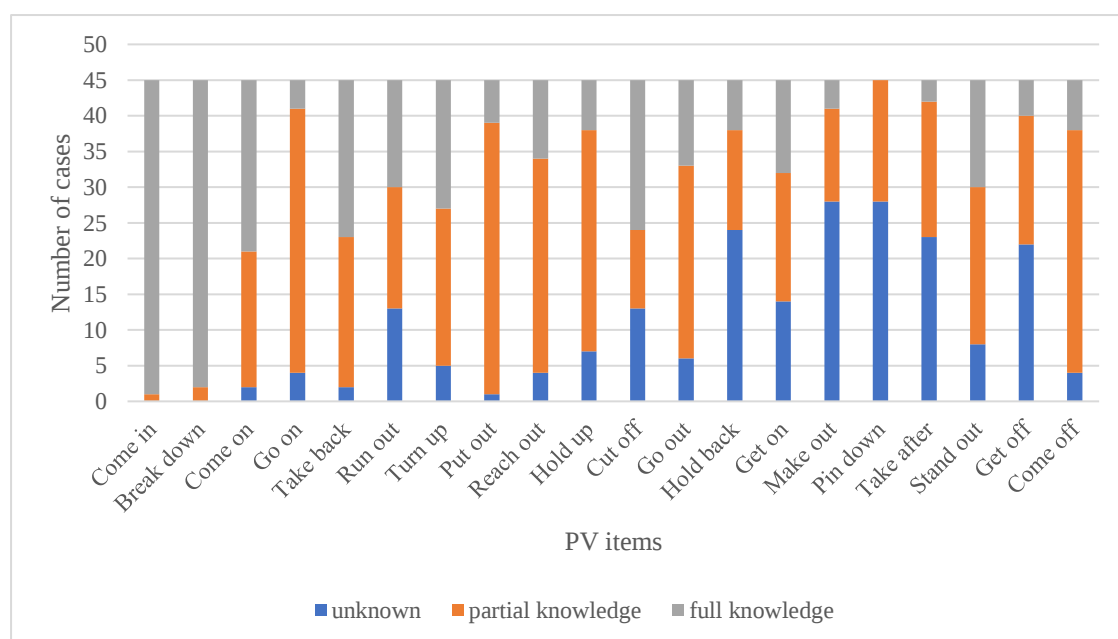
Figure 4.15 Coding of the developmental profile of PV knowledge

Gain	Stabilization	Attrition	Fluctuation
• unknown → partial knowledge • unknown → full knowledge • partial → full knowledge	• unknown → unknown • partial → partial knowledge • full → full knowledge	• full → partial knowledge • partial knowledge → unknown • full knowledge → unknown	• gain and attrition taking place between three time points

In the cross-sectional analysis of PV knowledge of tests at each time point (Figure 4.16), for each PV, there were 45 cases in total (15 participants * 3 times = 45). Most

participants (93% of a total N=15) had full knowledge (both productive and receptive command) of the tested meaning of *come in* and *break down*. Most (62.2%) of the sample could not recognize or use PVs such as *hold back*, *make out* and *pin down* (therefore coded as ‘unknown’). The majority (54.88%) of the sample had partial knowledge of the remaining 15 PVs.

Figure 4.16 Degree of knowledge of target PVs



In the longitudinal analysis of PV knowledge, the cases were calculated based on Figure 4.17. There were 149 cases (49.6%) of the same performance and 80 cases (26.7%) of gains over three time points. Additionally, there were 30 cases (10%) of worse performance at T3 and 41 cases (13.7%) of fluctuating scores (Table 4.12). In a breakdown of the types of gains, improvement in productive knowledge was more common than improvement in receptive knowledge and most gains took place between T1 and T2, rather than during the lockdown (Table 4.13). Lastly, regarding knowledge change on individual PVs, *hold back* saw the most cases of improvement, while *turn up* saw the most cases of attrition (Figure 4.17). *Come on* experienced the most cases of fluctuation. Except for *come in* and *break down* in which participants almost all got full scores, *put out* saw the largest number of stabilization cases, suggesting that their

error of using this PV did not change over one academic year.

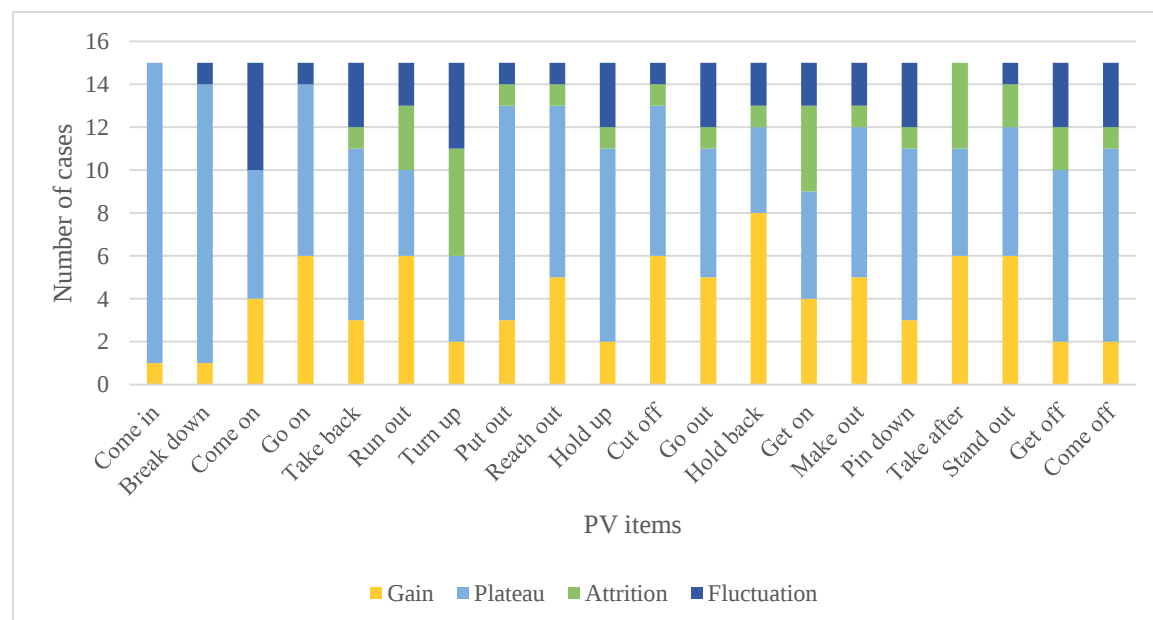
Table 4.12 Type of changes of PV knowledge (measured by numbers of cases)

Trajectory	Gain	Plateau	Attrition	Fluctuation
N	80	149	30	41

Table 4.13 Types of gains of PV knowledge (measured by numbers of cases)

Period	Partial-Full knowledge	Unknown-Partial knowledge	Unknown-Full knowledge	Total gains
T1-T2	33	25	4	62
T1-T3	44	29	7	80

Figure 4.17 Individual changing trends of PV knowledge



4.1.4.2 Error analysis of the receptive test

The errors made by each student that participated in the receptive test three times were examined in detail, in search of a systematic pattern of errors. These error analyses contribute indirectly to answering RQ1 because it reflects on the erroneous forms of PV which hindered their acquisition of correct PV forms. In the productive tests, respondents had different test-taking strategies and some people might not attempt the questions that they did not know and some people might put in words that were not

verbs or particles. Therefore, it is less justifiable to analyze errors in the productive tests.

In the receptive test, to find out if any systematic errors were made, erroneous PV structures (e.g., misuse of the verb or the adverbial particle) occupying more than 50% of all the erroneous combinations were recorded, as the erroneous form reached a majority in all wrong answers. It was found that four items met this condition. In Item 4 *The fire has finally been _____*. 82.9% of the non-target answers were *put off*. In Item 7 *Take the bus and _____ the bus at the third stop*. 63.6% of the erroneous answers referred to *go off*. In Item 15 *I think I _____ my mother. We look very similar and we like the same kinds of things*. 57.1% of the respondents answered *take along*. Lastly, for Item 18 *When we first met, we didn't like each other but now we _____ really well*. 75% of the respondents' non-target answer was *get over*. It can be inferred that students either had some vague impression of the right forms of PVs or they speculated the right verbs based on the context and the meaning. Particles were confusing for them because they were not certain enough about which particle to choose, due to limited familiarity with the target forms.

In brief, the above analyses answered RQ1 in two ways. Firstly, students' command of individual PVs was relatively stable over the one academic year and their knowledge of the majority of PVs remained the same. There were more gains than attrition on the tested PVs. The gains usually appeared incrementally, from the receptive state to the productive state, rather than from unknown to full knowledge. Interestingly, some students' knowledge fluctuated, showing that formulaic language acquisition is a reiterative process. Secondly, error analysis of the receptive test showed that one challenge that participants commonly faced was the adverbial particle use in PVs. It was found that the largest percentage of erroneous forms deviated from the target PVs in the particles. There were also other random mistakes on similar lexical verbs, but that did not form a majority of the errors.

4.2 LEARNER CORPUS

19, 21, and 16 participants completed the IELTS speaking test at T1, T2, and T3 respectively. 24 interviewees completed at least one IELTS speaking test, among which 10 completed it all three times with the researcher. The PV production by each interviewee from the learner corpus is presented in Table 4.14. The details of the speaking tasks can be seen in Table 3.12 in Section 3.4.4. Due to participants' different English oral fluency, after controlling the maximum test time (15 minutes), some low-proficiency participants only had time to answer one topic in Part 1, while the more fluent participants had time to answer two topics in Part 1. Thus, these less fluent students were asked fewer questions in Part 3. Since this time control was the same as the authentic IELTS speaking test, it was deemed fair for all participants. A list of all the PVs produced in each speaking test is provided in Appendix H.

In this section, the quantitative findings from the learner corpus are presented from three aspects, being the number of PVs produced (operationalized as types and tokens of PVs), the sophistication of PVs (defined as a person's command of less common PVs in native speakers' corpora) and the accuracy of PVs (defined as the ability to produce error-free PVs in the speech). In the SA research, errors in learner speech generally include both grammatical and lexical deviations from native-speaker norms (Leonard & Shea, 2017). Therefore, the accuracy measure was examined from the semantic accuracy of PVs and the grammatical accuracy of PVs. In particular, these variables between different time points were compared, to answer RQ1 on the developmental profile of PV command of learners. Next, some potential non-language-contact factors influencing the PV production in the learner corpus were examined.

Table 4.14 Overview of PV type number in the learner corpus

Student ID	T1P1	T1P2	T1P3	T2P1	T2P2	T2P3	T3P1	T3P2	T3P3
8111808	1	1	2	0	1	0	2	1	0
8109861	2	0	0	2	0	0	1	0	2
8111264				2	0	0	1	1	0
8111469	3	0	1	0	0	0	3	0	0

8112422	0	0	1				1	0	1
8111549				1	0	1			
8112158	1	0	0	3	0	0			
8110442	2	0	1	2	2	0	3	1	3
8110801	1	0	1	0	2	0			
8048132				0	1	0	3	1	4
8112153	0	0	1	0	0	0	0	0	0
8110416	0	0	0	0	0	0	0	1	0
8111871	1	0	4						
8109948	2	0	0	2	1	4	2	1	0
8110376	1	0	1						
8112165				1	1	1	1	1	0
8109688	1	1	2	1	0	0			
8112205				2	0	0	4	1	0
8109478	1	0	1	1	0	1	1	1	1
8113041	2	0	1	1	0	N/A			
8111535	5	1	3	3	0	N/A			
8110967	3	1	2	1	0	N/A	2	0	2
8048684	5	1	N/A	2	2	N/A			
8110193	2	0	1	2	0	2			
Mean	1.74	0.26	1.22	1.24	0.48	0.53	1.71	0.64	0.93
SD	1.41	0.44	1.03	0.97	0.73	1.04	1.16	0.48	1.28

Note: T=Time, P=Part in IELTS speaking test

N/A means these participants answered some Part 1 questions in Part 3.

4.2.1 Quantitative analysis of learner corpus

4.2.1.1 Number of phrasal verbs

In total, 56 PVs appeared 205 times in the learner corpus, based on the definitions from Oxford Learners' Dictionary and Oxford Learners' Phrasal Verb Dictionary (see Section 3.6.1.2 for dictionary selection). PVs with adverbial particles were counted, while PVs with prepositional particles were not counted (e.g., *belong to*, *go for*). This decision was made following the definition by Gardner and Davies (2007), for consistency with the current PV tests. The most common ten PVs from the learner corpus are presented below (Table 4.15). Because at T3 the number of interviewees reduced, the total number of PVs dropped accordingly. Thus, the frequency of the most common PVs at each time point had no comparison value. In total, *go out* (37), *focus*

on (18), *grow up* (18), *go back* (15) and *get up* (12) had the highest numbers of token occurrences. The most common lexical verbs from all the PVs were *go* (8 PVs), *get* (5 PVs), and *come* (4 PVs). Other verbs appeared in fewer than 4 PVs.

Table 4.15 Most frequent PVs from the whole learner corpus

PV	Frequency
go out	37
focus on	18
grow up	18
go back	15
get up	12
depend on	11
stay up	8
hang out	7
put in	5
give up	4

A range of values from the learner corpus data were calculated: the raw and normalized total tokens (word count), the raw and normalized PV types (each unique PV counted once), and PV tokens (including the same PV appearing multiple times). It was found that the percentage of PVs in total tokens was quite stable over three time points, being 0.5%, 0.4%, and 0.6% of the total tokens respectively. Taking the mean value of 0.5% as a usual appearance rate, it means that one PV appeared about every 200 words in the participants' English utterances, which is very close to the PV frequency that a learner might encounter (one in every 192 words) in naturalistic L1 English use, as extrapolated by Gardner and Davies (2007).

Table 4.16 Descriptive statistics of PV number

Variables	T1total tokens	T1 PV tokens	T1 PV types	T2total tokens	T2 PV tokens	T2 PV types	T3total tokens	T3 PV tokens	T3 PV types
Mean	694.58	3.84	2.89	672.14	2.62	2.05	699.07	4.29	3.00
SD	166.00	2.83	2.11	150.37	2.13	1.50	136.91	3.31	1.92
N	19	19	19	21	21	21	14	14	14

According to Table 4.16, at T1, among an average text length of 694.58 words, each participant on average produced 2.89 PVs 3.84 times. At T2, from the transcripts of an average length of 672.14 words, they produced 2.05 PVs 2.62 times. At T3, they spoke 699.07 words on average and produced 3 PVs 4.29 times on average. Overall, this indicates that descriptively there was a corresponding trend that the more words the interviewees produced, the more PVs they produced. Since T1 and T2 variables were normally distributed and T3 variables were skewed to the lower half, Pearson correlation and Spearman correlation between these key variables over three time points were performed to confirm whether a positive correlation existed.

It was found that from T2 onwards, there was a strong positive correlation between total tokens, PV types, and PV tokens. The coefficients in Pearson correlation and Spearman correlation were similar, so only the significant Pearson correlation results are presented here and that of the Spearman correction are attached in Appendix H. After Bonferroni correction, the significant threshold was $p=.008$ ($.05/6$), so only two pairs of significant correlations were maintained. More total tokens at T2 were associated with more PV tokens ($r=.570$, $p=.007$). Likewise, more total tokens produced at T3 were related to more PV tokens ($r=.685$, $p=.007$).

To find out if learners made gains in PV number, the normalized and the raw PV types and tokens were compared between time points. Since T3 variables did not meet normal distribution, parametric tests and their non-parametric counterparts were carried out. Paired-sample *t*-tests between T1 and T2 were employed to counterbalance the sample size decrease at T3, to make full use of the available sample. There was no significant gain after Bonferroni correction. When comparing data between three data points, there was no significant difference in PV types and PV tokens based on repeated-measures ANOVA and Friedman's Test ($p > .05$) (Appendix H).

Finally, the developmental profiles of PV number of individual learners are presented in the line graphs of PV types below. Since the raw PV types and the normalized PV type patterns are the same, only the normalized PV type patterns are displayed here

(Figure 4.18a, Figure 4.18b). Over three time points, there was no clear trend of change. On one hand, most participants' PV production was stable (almost horizontal line) (N=6). On the other hand, some participants' changes in PV number were dramatic, going in opposite directions or having steep slopes (N=3).

Figure 4.18a PV number change between T1 and T2 by participant (N=15)

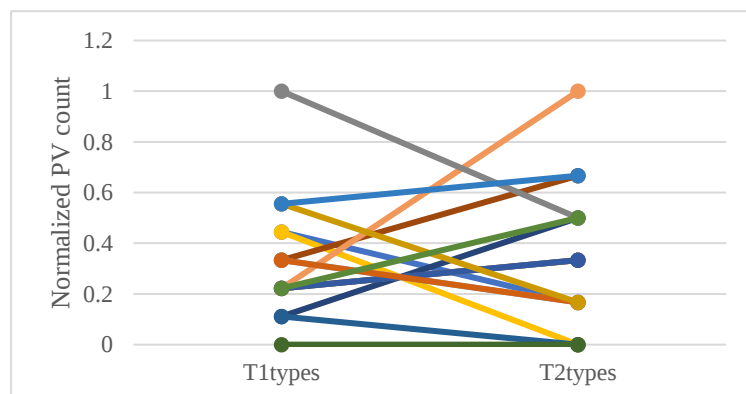
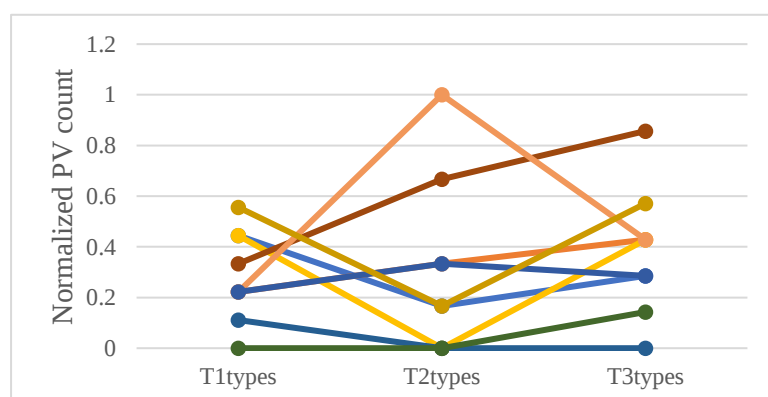


Figure 4.18b PV number change over T1, T2, and T3 by participant (N=9)



4.2.1.2 Sophistication of phrasal verbs

Lexical sophistication is the use of advanced words by learners (Failasofah & Alkhrisheh, 2018). However, the sophistication of PVs was hardly researched in SLA, so it was not well defined. Thus, the criteria to evaluate PV sophistication in this learner corpus was the feature of low-frequency of appearance in native speaker corpora (e.g., BNC) alone. The within-list PVs were coded based on the protocols in Table 3.13. For the 23 off-list PVs, their sophistication was doublechecked in the [BNC](#), rather than coded 4 directly. They were recoded 1 for 1000+ occurrences in the BNC (e.g., *focus*

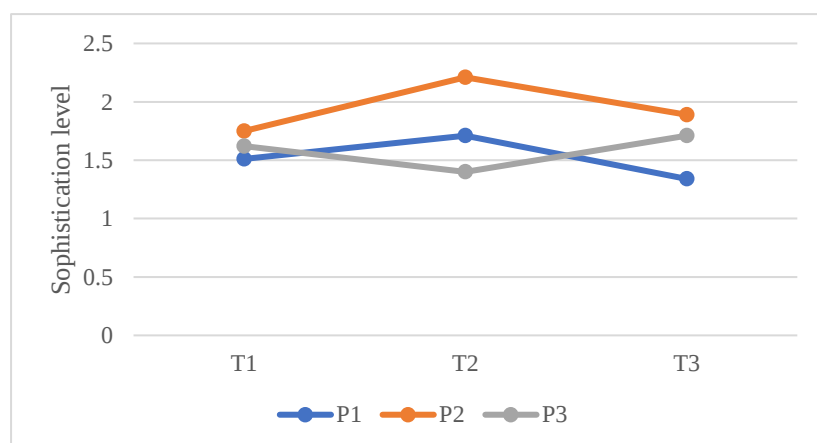
on, depend on), 2 for 501-1000 occurrences (e.g., *cut down*), 3 for 101-500 occurrences (e.g., *calm down, hurry up*), and 4 for less than 100 occurrences (e.g., *die out, pass away*). After comparing the corpus frequencies of some within-list PVs, their frequencies of presence were deemed comparable.

The mean PV sophistication at T1, T2, and T3 is shown in Table 4.17, suggesting a consistently low sophistication level (i.e., high corpus frequency). Wilcoxon Signed-Rank Tests and Friedman's Test were carried out because of the non-parametric nature of this variable. The results did not detect any significant difference between any time point ($p > .05$), suggesting this feature was stable over seven months (see Appendix H). The PV sophistication of each speaking part at each time point is presented in the line graph below to demonstrate the developmental profile of this variable (Figure 4.19). Eyeballing this line graph, between different speaking parts, Part 2 elicited the most sophisticated PVs because the participants had to provide rich details of certain experiences. Part 1 tended to bring about the lowest PV sophistication at each time point, which makes sense due to its simple conversation nature.

Table 4.17 Descriptive of sophistication of PVs in the learner corpus

Sophistication	T1	T2	T3
N	18	18	13
Mean	1.57	1.79	1.59
SD	0.65	0.74	0.44

Figure 4.19 Developmental profile of PV sophistication of each part



Note: P = speaking part

4.2.1.3 Accuracy of phrasal verbs

The accuracy of PVs in the learner corpus was examined in two ways, semantic and grammatical accuracy. The constructs were operationalized as the appropriate use of PVs for meaningful communication and the use of grammatically acceptable sentences respectively (Ragini, 2019). The findings are presented quantitatively with statistical analysis and qualitatively with erroneous examples. The overall mean semantic accuracy was .94 (SD =.19), whereas the overall mean of grammatical accuracy was .79 (SD =.39).

The descriptive statistics for each part of the speaking test are listed in Table 4.18a and Table 4.18b. The results show that the participants mostly used correct PVs to express their ideas but some failed to use the correct grammatical forms. Because these two variables were non-normally distributed with most values close to 1, which was not suitable for parametric tests, Wilcoxon Signed-Rank Tests and Friedman's Test were carried out instead. The results did not detect any significant difference between any time point ($p>.05$), suggesting this feature was stable over one academic year (see Appendix H for more details).

Table 4.18a Descriptives of semantic accuracy of PVs in the learner corpus

	T1	T2	T3
N	18	18	13
Mean	0.91	0.94	0.98
SD	0.25	0.16	0.07

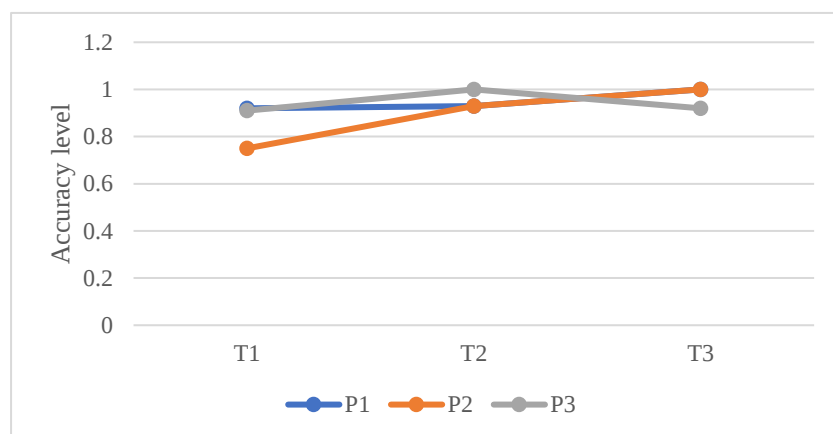
Table 4.18b Descriptives of grammatical accuracy of PVs in the learner corpus

	T1	T2	T3
N	18	18	13
Mean	0.60	0.70	0.82
SD	0.36	0.34	0.21

The line graphs of semantic accuracy and grammatical accuracy changes are presented in Figure 4.20a and Figure 4.20b to demonstrate their developmental profile. Between

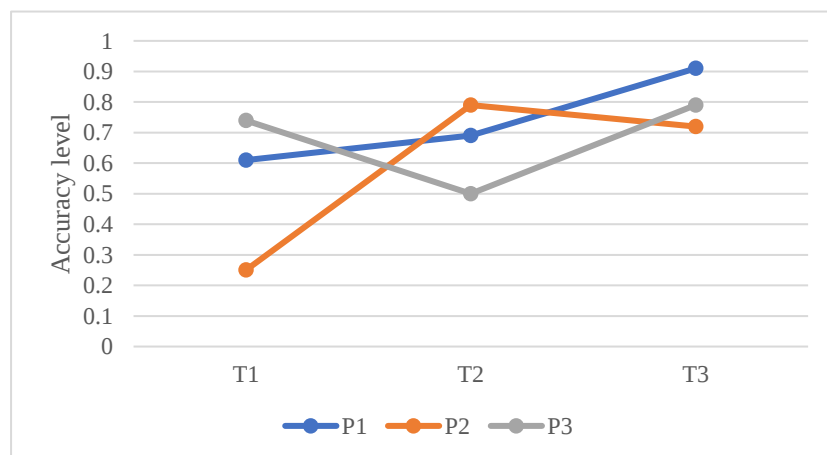
time points, semantic accuracy plateaued without much variation, whereas the grammatical accuracy witnessed some fluctuation. Regarding speaking parts, in Part 2, participants had some trouble producing correct PVs semantically and grammatically (the orange line), perhaps because of the small sample size and the challenges of the monologue task itself. To be specific, smaller sample size was more likely to suffer from outliers and the limited cognitive resources of participants could not be devoted to meaning and grammar at the same time. For example, only four interviewees used PVs in Part 2 at T1. The topic was about a memorable trip, which had the lowest grammatical accuracy of PVs.

Figure 4.20a Developmental profile of PV semantic accuracy of each part



Note: P = speaking part

Figure 4.20b Developmental profile of PV grammatical accuracy of each part



Note: P = speaking part

4.2.1.4 Summary of quantitative analysis of phrasal verbs

Overall, the token of PVs produced in the learner corpus was significantly related to the total words produced in the IELTS speaking test. The sophistication of the PVs was consistently low between three time points, showing that participants might only automatize the high-frequency PVs. The semantic accuracy of the produced PVs was higher than their grammatical accuracy, suggesting that the sample focused on meaning expression more than accuracy. The types and tokens, sophistication, and semantic and grammatical accuracy of PVs did not see significant growth from T1 to T3.

4.2.2 Qualitative analysis of inaccurate PV use

In the qualitative analysis of these data, the PVs with semantic errors and the PVs with grammatical errors are presented with typical examples. Regarding semantic errors, the errors were categorized into two types (Table 4.19). The respondents tended to over-extend the meaning of some common PVs. For example, in T1 Part 1, *cut down* was used improperly in **I think I must go to China Town to cut down my hair.* (AC15) At T1 Part 2, Interviewee AC9 said **I have never go out in the other country except China.* At T1 Part 3, *come in* was mistakenly used to refer to ‘appear’ in the sentence **When they are older, the body warning will come in.* (AC18) Other errors indicated a lack of collocation knowledge of some PVs. For example, **Some new transportation will come up.* (AC9) **Some resources will (is) not coming up, so we need to save them.* (AC9) In this context, the first *come up* should be replaced with ‘appear’ and the second *come up* should be replaced by ‘renew’.

Table 4.19 Coding scheme for semantic errors

Code	Frequency	Example
Wrong meaning	6	*When they are older, the body warning will <u>come in</u> .
Wrong collocation	4	*If you <u>take up</u> your head, you can always see them.

In terms of grammatical accuracy, participants had lower mean scores. There were four

types of errors (Table 4.20). The most common problem was that participants used the present simple tense to describe actions in the past, which affected their grammatical accuracy. For example, when describing a memorable trip in Part 2 at Time 1, Interviewee AC3 said **In the morning we get up really late and have a good rest.* Secondly, students were unfamiliar with some necessary participle changes of the lexical verbs in a PV (to do or doing). For instance, **I like walk around.* (AC18) **After staying alone too much, you prefer go out with your friend.* (AC18) It is a convention that present participles (verb + ing form) are used after some particular verbs (e.g., like, prefer) and participants were unaware of that. Thirdly, the use of gerunds was not mastered. For example, **the school can require students to do things to keep healthy, for example get up early and go to bed early.* (BC5) **Except for I stayed up very late, I would get up at 10.* (BC5) There were clear signs (e.g., for example, except for) that gerund should be used. These errors reflected the limited grammatical competence of these participants and the errors did not pertain to PVs alone: they were also common in the non-PV production in the speaking data. Thus, the grammar knowledge of other verb usage may influence the grammatical accuracy rating of the PV use by the participants.

Table 4.20 Coding scheme for grammatical errors

Code	Frequency	Example
Tense	18	<i>*In the morning we get up really late and have a good rest.</i>
Verb + doing	8	<i>*I always like sleep late and get up late.</i>
Gerund	2	<i>*Except for I stayed up very late, I would get up at 10.</i>
To do	1	<i>*The height of the mountain is high enough for this mountain to went over the clouds.</i>

4.2.3 Non-exposure factors that influence phrasal verb production

In this section, potential non-language-contact factors influencing the PV number, sophistication, and accuracy in the learner corpus were analyzed from the learner perspective (e.g., PV test scores, overall English proficiency) and the speaking test

perspective (e.g., topic type, task type). These factors were selected because higher proficiency of learners is usually associated with better vocabulary command (Schmitt, 2014) and the use of PVs is heavily influenced by register (e.g., spoken vs written, newspaper vs fiction) (Liu, 2011). But little is known whether the topic type or task type of a speaking test influences PV characteristics. Different techniques were used because some variables were normally distributed (e.g., raw PV number) while some were not (e.g., PV sophistication, grammatical and semantic accuracy). A summary of results is displayed in Table 4.21 and none of these factors were significant after Bonferroni correction (.005=.05/10).

Table 4.21 Summary of factors investigated and statistical techniques

Factor	Techniques	Significance	Corrected significance
PV test scores	Correlation	Yes	No
Overall English proficiency	Correlation	Yes	No
Speaking test topic types	One-way between-groups ANOVA, Kruskal-Wallis test	No	No
Speaking test task types	Independent-samples t-test, Mann Whitney U test	Yes	No

4.2.3.1 PV test scores

In order to find out the relationships between PV knowledge in the controlled measures and that in uncontrolled measures, the PV productive and receptive test scores at T1, T2, and T3 were correlated with variables of PV characteristics (number, sophistication, semantic accuracy, grammatical accuracy) from the learner corpus by Pearson correlation and Spearman correlation, depending on normality of distribution. Details of the correlation results are displayed in Appendix H.

4.2.3.2 English proficiency

In the current study, the self-reported IELTS score at T1 and their VLT scores were two proxies of students' initial and overall English proficiency. In the parametric and non-parametric correlation analysis between PV type, accuracy, and sophistication in the

learner corpus and initial IELTS scores/VLT scores at T1 and T2, the positive correlations were not significant after Bonferroni correction. Details of the correlation results are displayed in Appendix H.

4.2.3.3 Speaking topics

Speaking topics might have a direct impact on the number of PVs that the interviewee used in the speaking test, as some topics entailed rich descriptions of daily routines or actions. In the current IELTS speaking test, the researcher purposely chose topics that are likely to elicit PVs based on piloting and discarded daily routine questions which definitely entail PVs like *wake up*, *get up*. All the speaking topics were further categorized into three types, being descriptive topics (N=4, e.g., your study, your hometown, your families and friends, a film), lifestyle topics (N=5, e.g., evening routines, free time, hobbies, healthy lifestyle, leisure activities), and travel topics (N=3, e.g., a memorable trip, means of transportation). A between-groups ANOVA and the Kruskal-Wallis test revealed that the PV number, sophistication, semantic accuracy, and grammatical accuracy did not differ significantly between these three topics ($p > .05$) (see Appendix H for more details). This result suggested the selection of speaking topics was fair and the PV eliciting power of these topics was similar.

4.2.3.4 Task types

Finally, the role of task type (dialogue vs. monologue) was also investigated. The speaking parts were labeled as either dialogue (Part 1 and Part 3) or monologue (Part 2). The significant results in independent-samples *t*-tests and Mann-Whitney U tests were the same (Appendix H). After Bonferroni correction ($p = .0125 = .05/4$), the differences were non-significant.

4.2.3.5 Summary of factors influencing PVs in the learner corpus

The analyses above have shown that learner factors and speaking test features had some neglectable relationships with PV production in the IELTS speaking test, despite the

small sample size at each time point. Regarding learner factors, higher PV test scores and higher IELTS scores were positively correlated with higher PV number, sophistication, semantic and grammatical accuracy in the learner corpus at different time points. In terms of task features, the chosen topics did not have an evident impact on the PV production. The dialogue task type produced more PVs, whereas the monologue task type led to higher PV sophistication. However, these associations dissipated after Bonferroni correction was applied, showing their weak relationships.

4.3 SUMMARY OF PHRASAL VERB DEVELOPMENT

RQ 1 focuses on the developmental profile of PV knowledge of international foundation year students from November 2019 to May 2020. Students' performance on PV tests and PV learner corpus showed a different pattern. In the PV tests, significant gains were observed, especially in the productive tests. The gains were recorded between T1 to T2 and T1 to T3. The gain in the receptive test was smaller and was only documented between T1 to T2. In the qualitative analysis of the developmental trajectory of PVs in both tests, it was found that for most PVs, students' performance was stable over three time points. The cases with a trajectory of gain were more than the cases with trajectories of fluctuation or attrition, showing that students did increase their PV knowledge to some extent. However, their production of PVs in the IELTS speaking test was mostly stagnated. The participants tended to produce highly common PVs with the right meaning but with problematic grammar. No significant gain was directly captured in terms of PV number, PV sophistication, and PV accuracy. It can be speculated that changing the speaking preference of learners might take a long time or more intense L2 exposure.

As for RQ 1a, a number of non-language-contact factors were examined against students' PV performance in the test and the speech samples. It was found that the overall English proficiency of learners and corpus frequency of PVs had strong relationships with the PV test scores. To be specific, the higher English level of learners

and high corpus frequency of PVs were associated with higher PV scores. On the other hand, the role of semantic opacity and having Chinese as L1 were not confirmed as significant in the present study. In terms of the learner corpus, more total tokens (words) of the speech samples were strongly associated with more PVs produced. The role of PV test scores and overall English proficiency of learners, speaking topics, and task types were not detected as significant from the current learner corpus. A larger learner corpus is needed to further probe the impact of these factors on the sample.

CHAPTER V FINDINGS PART II

INFORMAL LANGUAGE CONTACT

This finding chapter aims to answer a main research question (RQ) and two sub-questions of this study:

RQ2. What are the patterns and developmental profile of informal L2 contact among international foundation program students in the UK for the first six months of the foundation program?

2a. What personal background factors influence the informal L2 contact of the participants?

2b. How did the COVID-19 lockdown in the UK in Spring 2020 affect participants' SA experience and informal L2 contact?

This chapter is divided into two parts based on the RQs. The pre-COVID section consists of answers to RQ2 and RQ2a, while the post-COVID section address RQ2b. In each part, data yielded from three instruments were used: the quantitative data from the Language Contact Questionnaire (LCQ), quantitative data from the Study Abroad Social Network Questionnaires (SASNS), and qualitative data from the semi-structured interviews. While the quantitative data provide a holistic picture of L2 use patterns of the whole sample, the qualitative data supplement the quantitative data in explaining the nuances of language contact differences between individuals and the L2 use not captured by close-ended questionnaires. Although the outbreak of COVID-19 was not anticipated by the researcher, this chapter offers an invaluable contrasting view on how the participants' SA experience was fundamentally changed by this pandemic.

5.1 PRE-COVID LANGUAGE CONTACT

In this section, findings from the LCQ, the SASNS, and the interviews completed at T1 (November 2019) and T2 (February 2019) are presented to illustrate a comprehensive language contact pattern of the sample. The LCQ focuses on the quantity and types of both interactive and non-interactive L2 contact, whereas the SASNS focuses on the quality of interactive L2 contact. Since students had very limited English-speaking social networks, especially at T1, the completion rate of SASNS was lower than LCQ.

5.1.1 The Language Contact Questionnaire

The LCQ will be analyzed from four aspects, being descriptive statistics, comparison of means, exploratory factor analysis (EFA), and potential influencing factors. The normality check was done in the first place for the selection of parametric/non-parametric analysis. The distribution of normality was checked in a few ways, ranging from the Kolmogorov-Smirnov test, Shapiro-Wilk test, Normal Q-Q Plot, the boxplot, the 5% trimmed mean, to eyeballing the histogram (Tabachnick & Fidell, 2013). The Cronbach's α of the LCQ L2 items at three time points were .881, .884, and .915 respectively, demonstrating a high level of internal reliability.

5.1.1.1 Descriptives and normality

5.1.1.1a Time 1 and Time 2 data

At T1, 175 valid LCQ responses were analyzed, with students' responses coded 0 to 5 based on frequency from "never" to "all the time". The possible maximum score for the LCQ was 80. The overall LCQ score mean was 42.02 (SD=11.27). The maximum score in the data was 70, whereas the minimum score was 19. Based on the normality check, this variable reached good normality (Figure 5.1). At T2, 153 valid LCQ responses were analyzed. The overall LCQ scores had a mean of 39.69 (SD=11.55). The maximum score was 66, whereas the minimum score was 8. The variable reached

good normality (Figure 5.2). Two low-scoring outliers were kept in the dataset since these responses were genuine (response for each item not identical) and they participated in both rounds of data collection.

Figure 5.1 Histogram of T1 LCQ scores

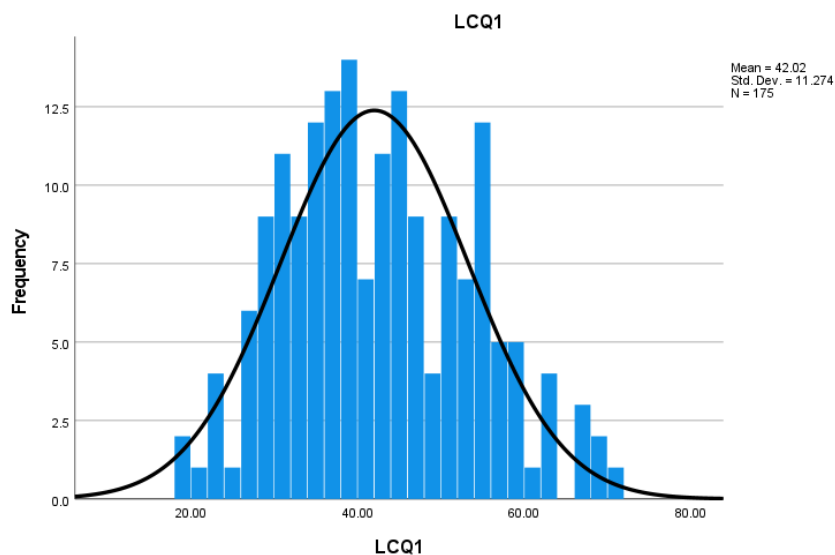
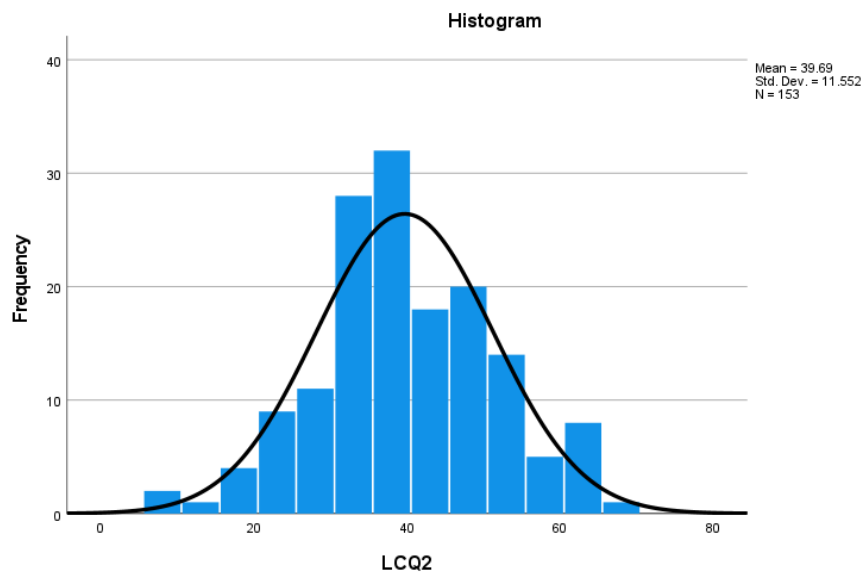


Figure 5.2 Histogram of T2 LCQ scores



In terms of individual LCQ items, the mean score of all L2 contact at T1 was 2.96 out of a possible maximum of 5, approximating the frequency of ‘often’ (Table 5.1). The items with the highest frequency were writing homework in English, speaking English with service personnel, speaking with L1 friends, and speaking English with

international friends. The participants spent the least time watching TV in English and reading books, newspapers, and magazines in English. At T2, the mean score of all L2 contact was 2.67, still approximating the category of ‘often’ (Table 5.1). The items with the highest contact were writing homework in English, speaking with L1 friends, using social media in L1, and speaking English with service personnel. The participants spent little time watching TV in English, playing computer games in English, reading books, newspapers, and magazines in English, and listening to radio or podcasts in English. The pattern implies that compulsory L2 use and L1 interactive use were more frequent than leisure L2 use among the participants.

Table 5.1 Descriptive statistics of T1 and T2 LCQ breakdown

English-contact items	T1 (N=175)		T2 (N=153)	
	Mean	SD	Mean	SD
Writing homework	4.04	0.99	3.91	1.20
Speaking with service personnel	3.78	1.27	3.4	1.34
Speaking with L1 friends*	3.68	1.26	3.87	1.30
Speaking with international friends	3.55	1.34	2.94	1.48
Listening to songs	3.45	1.24	3.26	1.41
Using social media (e.g., Facebook, WhatsApp)	3.14	1.49	2.78	1.64
Writing Emails	3.02	1.25	2.95	1.19
Reading timetables, menus, maps, announcements, etc.	3.01	1.14	2.97	1.29
Listening to people’s conversation	2.99	1.13	2.94	1.17
Watching movies	2.93	1.18	2.76	1.31
Speaking with British friends	2.93	1.79	2.53	1.61
Using social media in L1*	2.91	1.56	3.53	1.48
Watching YouTube or Netflix	2.73	1.35	2.56	1.34
Browsing websites	2.50	1.22	2.56	1.21
Playing computer or mobile games	2.31	1.61	1.99	1.64
Watching TV	2.18	1.34	2.21	1.38
Reading books, newspapers, or magazines	1.90	1.12	1.66	1.06

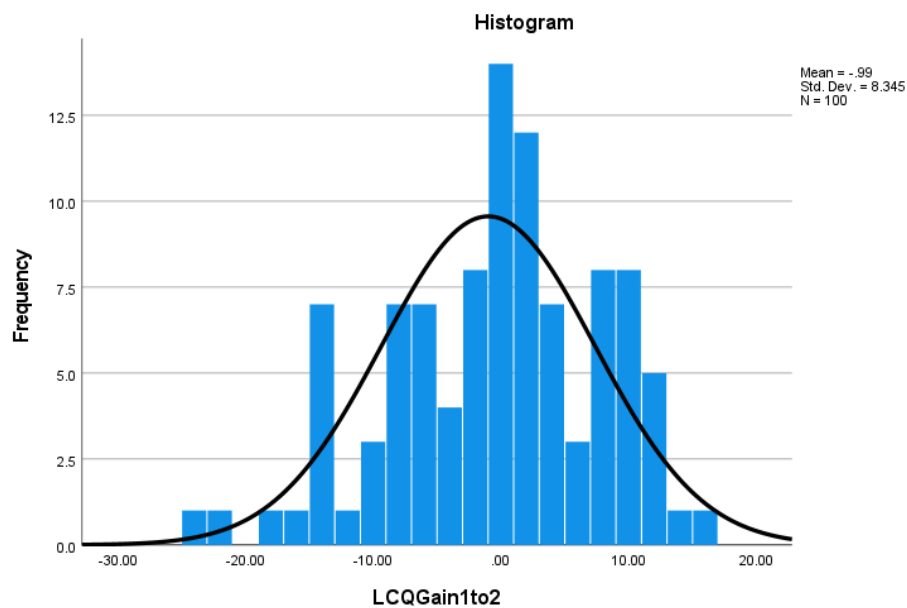
Note: *L1-contact items

5.1.1.1b T1 to T2 gain

The L2 contact gain was measured by the LCQ scores at T2 subtracting that at T1. The descriptives of T1-T2 overall LCQ gains and individual item changes are displayed

below. The overall LCQ T1-T2 gain had a mean of $-.99$ ($SD=8.35$, $N=100$). The maximum gain was 15, whereas the minimum gain was -24. The distribution of normality was checked in the abovementioned ways (Figure 5.3). Although there were two low-scoring cases, their responses were kept because they completed the instrument twice and unengaged responses (identical answers) were not spotted.

Figure 5.3 Histogram of T1 to T2 LCQ gains



With regards to individual item changes, the mean score change was $-.01$ out of a possible range of 5 and -5 (Table 5.2). The items with the highest frequency gains were using social media in L1, speaking with L1 friends, and writing Emails in English. The items that witnessed the largest frequency decrease were speaking with international friends in English, speaking with service personnel in English, using social media in English, and speaking with British friends in English. It could be seen that an upward trend took place in L1 activities and a downward trend was observed in L2 socialization.

Table 5.2 Descriptive statistics of the T1 – T2 LCQ item gains

English-contact items	Mean	SD	Min	Max
Using social media in L1*	0.54	1.47	-4	4
Speaking with L1 friends*	0.17	1.51	-3	5
Writing Emails	0.16	1.39	-4	4
Watching TV	0.16	1.29	-3	4

Listening to people's conversation	0.15	1.26	-5	3
Browsing websites	0.13	1.13	-2	3
Reading timetables, menus, maps, announcements	0.09	1.09	-3	3
Watching YouTube or Netflix	0.01	1.47	-4	4
Writing homework	0.00	1.06	-3	2
Listening to songs	-0.01	0.95	-2	3
Watching movies	-0.02	1.14	-3	3
Reading books, newspapers or magazines	-0.10	0.73	-2	3
Speaking with British friends	-0.28	1.85	-5	5
Using social media	-0.34	1.55	-4	5
Speaking with service personnel	-0.39	1.43	-4	3
Speaking with international friends	-0.45	1.40	-5	3

Note: *L1-contact items

5.1.1.2 Comparison of means

Paired-samples *t*-tests were performed to find out whether significant changes took place at the total score and item level during the first six months of the normal SA period (i.e., between T1 and T2). Like in the previous chapter, Bonferroni correction was adopted to minimize the risk of committing Type I errors. It should be noted that the total LCQ scores always excluded the two L1-using items, while in item-level comparison the two L1-using items were included.

Paired-samples *t*-test revealed that there was no statistically significant gain in the overall LCQ scores from T1 ($M=40.12$, $SD=10.07$) to T2 ($M=39.03$, $SD=12.18$): $t(99) = -1.289$, $p = .200$, $N = 100$. In terms of individual LCQ items, two changes were statistically significant (Table 5.3). The frequency of using social media in L1 at T1 increased at T2. The mean gain was .54, with a large effect size. On the contrary, the frequency of speaking English with international friends lowered at T2. The mean decrease was -.46, with a medium effect size. The non-significant frequency declines of other English-using items are shown in Appendix I. Overall, this result suggested that although the overall L2-using frequency remained stable, the sample steered toward their L1 in their leisure time, rather than speaking English more.

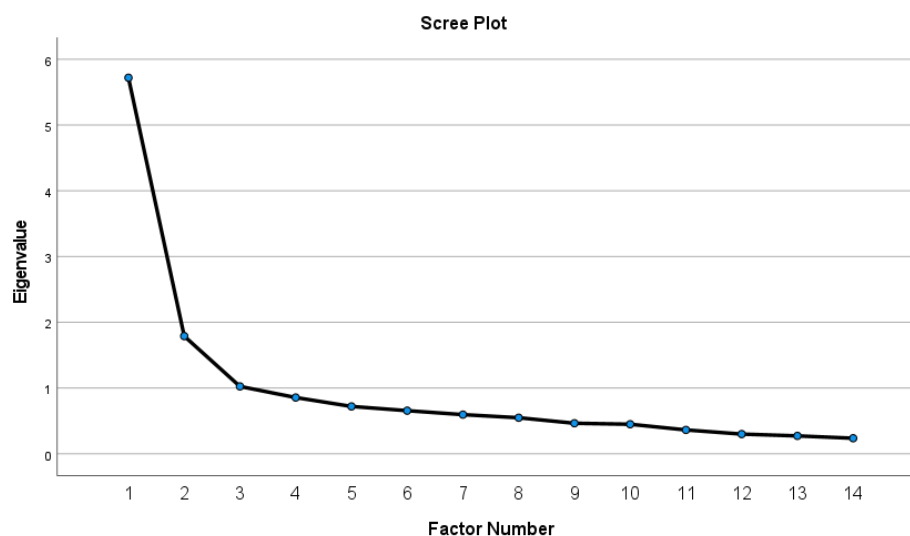
Table 5.3 Comparison of LCQ items between T1 and T2

Item		Mean	SD	t	df	p	N	partial η^2
Using L1 social media	T1	3.00	1.44	3.666	99	.000	100	.12
	T2	3.54	1.49					
Speaking with international friends	T1	3.40	1.41	-3.227	98	.000	99	.09
	T2	2.95	1.46					

5.1.1.3 Exploratory factor analysis

To understand the latent constructs underlying the L2 use items, EFA on T1 and T2 data were performed. A few assumptions were checked prior to running the EFA. Firstly, the sample size at T1 was 175 and that at T2 was 153 and there was no noticeable outlier, which allowed for EFA. According to Tabachnick and Fidell (2013), a 10:1 case item ratio should be sufficient for this analysis, so a sample size larger than 150 sufficed. The correlation matrix confirmed that there was an absence of multicollinearity and singularity. The factorability of the correlation matrix was confirmed by the significant p -value in the Bartlett's Test of Sphericity ($p < .001$) and the KMO value (.882) which was larger than .60. To decide the number of factors to retain, two methods were employed: the Kaiser's criterion (eigenvalue greater than one) and the scree plot trend (look for the elbow turn). Therefore, it was decided to keep three reproduced factors in the EFA (see Figure 5.4 for the T1 scree plot and see Appendix I for the T2 scree plot).

Figure 5.4 Scree plot of T1 LCQ EFA



At T1, because no variable's communalities extraction was smaller than 0.3, they were all kept for the EFA. Because the coefficient in the factor correlation matrix was larger than 0.2, the oblique rotation method was employed. Promax is the most common oblique rotation method deemed by much literature (J. D. Brown, 2009). With the extraction method of Principal Axis Factoring and the rotation method of Promax, the following three factors were identified (coefficient values lower than 0.4 were suppressed). As seen in Table 5.4, Factor 1 mainly encompasses receiving visual or audio English input, so it was labeled *receptive contact*. Factor 2 entails English spoken output activities, so it was labeled *speaking contact*. Finally, factor 3 only contains writing output activities, so it was labeled *writing contact*. Literature suggested that in a good dataset with a clear structure, the interpretations between different extraction and rotation methods should be the same (Tabachnick & Fidell, 2013). In this dataset, the pattern matrix using the Direct Oblimin rotation method resulted in highly identical factor loadings, so it is not presented here. The internal reliability of T1 reproduced factors were satisfactory (receptive contact: Cronbach's $\alpha = .858$; speaking contact: Cronbach's $\alpha = .721$; writing contact: Cronbach's $\alpha = .726$).

Table 5.4 Pattern Matrix of T1 LCQ data

English-using items	Factor 1 receptive	Factor 2 speaking	Factor 3 writing
Watching movie	.900		
Watching YouTube or Netflix	.812		
Watching TV	.750		
Listening to people's conversation	.590		
Using social media (e.g., Facebook, WhatsApp)	.560		
Browsing websites	.556		
Reading timetables, menus, maps, announcements, etc.	.528		
Listening to songs	.491		
Reading books, newspapers, and magazines	.476		
Speaking with British friends		.938	
Speaking with international friends		.751	
Speaking with service personnel		.506	
Writing homework			.881
Writing Emails			.625
Extraction Method: Principal Axis Factoring.			

Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

As for T2 EFA, the same procedure was run and the data yielded three factors (Table 5.5). Factor 1 consists of leisure activities receiving audio and visual English input, so it was labeled *multimedia contact*. Factor 2 comprises activities producing English verbally or in writing. Listening to people’s conversations also usually took place in communication. Therefore, this factor was labeled *productive contact*. Factor 3 only included two activities entailing reading dense texts in English, so it was labeled *reading contact*. The internal reliability (Cronbach’s α) of each factor was .800, .823, and .680. Each factor was intentionally labeled differently from the T1 factors to avoid confusion. It can be seen that the major differences between T1 and T2 EFA factors were that speaking contact and writing contact were merged into productive contact, while receptive contact was broken down into multimedia contact and reading contact. It is suggested that students’ frequency in reading English and enjoying multimedia recreation in English diverged at T2. While the majority of participants still watch a lot of English movies and TV series as pastimes, students who kept on reading in English are those who really felt comfortable and motivated in using the L2 out of class.

Table 5.5 Pattern Matrix of T2 LCQ data

English-using items	Factor 1 multimedia	Factor 2 productive	Factor 3 reading
Watching movie	.788		
Watching TV	.736		
Watching YouTube or Netflix	.728		
Listening to songs	.669		
Using social media (e.g., Facebook, WhatsApp)	.612		
Playing computer or mobile games	.498		
Speaking with service personnel		.739	
Writing homework		.619	
Speaking with British friends		.603	
Listening to people’s conversation		.594	
Writing Emails		.569	
Speaking with international friends		.563	

Reading timetables, menus, maps, announcements			
Reading books, newspapers, and magazines			.851
Browsing websites			.576
Listening to radio or podcast			
Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.a			
a. Rotation converged in 5 iterations.			

5.1.1.4 Factors influencing language contact

To address RQ2a, a number of factors that might influence the amount of language contact were explored in this section, being students' nationality, motivation, overall English proficiency, age, living arrangements, and center locations. These factors could affect their communication ability, intercultural awareness, and L2 using opportunities. Similar to the PV tests, these factors were analyzed to decide whether to include any of them in the final mixed-effects model.

5.1.1.4a Nationality

Due to the small number of non-Chinese participants, they were grouped into one subcategory, similar to the procedure of answering RQ1 on phrasal verbs. Independent-samples *t*-tests revealed that there were significant differences between the total amount of L2 contact between Chinese participants and non-Chinese participants at all time points with large effect sizes. The total LCQ scores of non-Chinese students were always higher than that of the Chinese students (Table 5.6). However, the LCQ change between different time points did not show a significant difference between Chinese and non-Chinese participants. Regarding language contact factors from the EFA, Chinese and non-Chinese students still demonstrated significant differences. The L2 receptive contact of Chinese students was significantly lower than that of non-Chinese students ($p < .001$), while the L2 speaking contact of Chinese participants was also lower than their non-Chinese counterparts ($p = .001$). Their amount of L2 writing contact did not differ significantly ($p = .072$). Lastly, in terms of individual LCQ items, significant differences between the Chinese and non-Chinese participants persisted. These two

subgroups saw significant differences in ten items at T1 and T2 respectively (see Appendix I). In short, the Chinese L1 was a strong independent factor influencing the language contact pattern of the participants.

Table 5.6 Significant nationality differences in LCQ scores

Item		Mean	SD	N	t	df	p	partial η^2
LCQ T1	Chinese	38.80	9.87	122	6.01	167	.000	.18
	Non-Chinese	49.38	11.19	47				
LCQ T2	Chinese	36.90	10.90	118	5.32	147	.000	.16
	Non-Chinese	48.65	11.06	31				
Receptive contact T1	Chinese	-.29	.79	122	73.39	98	.000	.34
	Non-Chinese	.67	.96	48				
Speaking contact T1	Chinese	-.17	.93	122	3.43	168	.001	.34
	Non-Chinese	.37	.88	48				

5.1.1.4b English proficiency

Pearson correlation was performed on the English proficiency and LCQ scores, to investigate possible relationships. The same as in Chapter Four, IELTS scores and VLT scores were selected as proxies of initial and overall English proficiency. On total LCQ scores, it was found that IELTS scores had a weak positive relationship with T1 LCQ scores and T2 LCQ scores (Table 5.7). T1 VLT scores were moderately correlated with T1 LCQ scores, whereas T2 VLT scores were weakly related to T2 LCQ scores.

Table 5.7 Correlation between overall English proficiency and LCQ scores

Item	IELTS			VLT1			VLT2		
	r	p	N	r	p	N	r	p	N
LCQ1	.223**	.012	126	.350**	.000	133	-	-	
LCQ2	.307**	.008	73	-	-		.281**	.001	141

** Significant after Bonferroni correction

Regarding the types of L2 contact identified in the EFA, IELTS scores only had a weak relationship with receptive contact ($r = .256$, $p = .004$, $N = 127$), but not speaking contact and writing contact. Likewise, T1 VLT scores had a moderate relationship with receptive contact ($r = .321$, $p < .001$, $N = 130$), but not speaking contact and writing contact. T2 VLT scores were also significantly correlated to multimedia contact (r

=.243, $p = .004$, $N = 135$) and reading contact ($r = .365$, $p < .001$, $N = 135$) at T2, but not productive contact. Lastly, in terms of individual L2-using activities, the frequency of information-intensive receptive activities (e.g., reading books, browsing websites) had a weak positive relationship with IELTS scores and VLT scores at T1 and T2 (Appendix I). Overall, the relationship between proficiency and L2 contact was generally weak and only limited to receptive contact. This result implied that students with a higher English level were more likely to read or watch contents in English out of class.

5.1.1.4c Motivation

The mean score of four Ideal L2-Self motivation items in the background questionnaire was computed and correlated with LCQ scores. It was found that motivation scores had weak to moderate relationships with the total LCQ scores and most LCQ factors at T1 and T2 (Table 5.8). It showed that highly motivated learners did have a higher frequency of informal L2 usage after class. It should be noted that motivation data was collected at T1, so its weaker relationship with T2 LCQ scores implied that this construct was highly dynamic and could fluctuate over time.

Table 5.8 Correlation between motivation and LCQ scores

T1 (N=175)			T2 items (N=100)		
Items	r	p	Items	r	p
LCQ1	.405**	.000	LCQ2	.271**	.006
Receptive contact	.387**	.000	Multimedia contact	.334**	.001
Speaking contact	.330**	.000	Productive contact	.206	.039
Writing contact	.184**	.015	Reading contact	.155	.121

** Significant after Bonferroni correction

5.1.1.4d Age

Correlation analysis was performed between age and T1 and T2 LCQ total scores and LCQ factors respectively. None of the correlation reveals significant relationships

(details can be found in Appendix I), implying that age did not influence the amount of time spent on using English among the sample.

5.1.1.4e Living arrangement

Participants in the sample mostly lived alone (N=82), lived with co-nationals (N=42), or lived with international friends (N=45). A between-groups ANOVA was performed on the T1 LCQ scores and three different living conditions, which showed significant results [$F(3, 168) = 3.527, p=.016$]. The post-hoc comparison showed that those who lived with international flatmates (M=46.73, SD=9.82) had significantly higher language contact than those who lived alone (M=40.94, SD=11.07, $p=.031$) or lived with co-national flatmates (M=40.00, SD=11.95, $p=.030$). No significant difference was captured in this regard when using LCQ factors, as the variance was too small after the EFA.

5.1.1.4f Center locations

Independent-samples *t*-test was performed on Center A students in London and Center B students in West Midlands. It was found that at T2 there was a significant difference between the language contact of participants in the two centers: Center A (M=41.54, SD=10.57; Center B: M=37.23, SD=12.87; $t(152) = 2.28, p=.024$), but not at T1. However, this significance might be explained because at Center B high-achievers were exempted from English classes at T2. Comparing the LCQ factors across locations at T1, with Bonferroni correction, receptive contact and speaking contact did not differ significantly. But writing contact at Center A (M=.22, SD=.81, N=89) was higher than that at Center B (M=-.23, SD=.89, N=87): $t(174) = 3.538, p=.001$. This could be related to different arrangements between the two centers. Center B had an enrolment partnership with University B. Center A students could apply to any university so they often wrote emails to admission offices of different universities for inquiry.

5.1.1.5 Summary of the LCQ data

To sum up, the abovementioned analysis showed that the overall LCQ scores at two time points met the normal distribution for parametric analysis. Generally speaking, L1 was frequently used in the social interaction of the sample. Further, their non-interactive L2 contact (e.g., writing homework) was more frequent than interactive L2 contact. The total L2 contact at T1 did not differ from T2 significantly, but a decreasing trend of L2 speaking contact was observed. The personal factors that influenced the LCQ were L1, motivation, and overall English proficiency. Chinese nationals had significantly less English contact and students with a higher English level had slightly more receptive L2 contact. Motivated students had more L2 contact in general. Those living with international flatmates had significantly higher L2 contact than those living by themselves or with L1-speaking friends. The center arrangement could have a slight impact on the amount of writing contact of the participants

5.1.2 The Study Abroad Social Network Questionnaire

The SASNS collected rich data on the friends' characteristics of the participants. In this section, the data were analyzed from a few aspects. Prior to the analysis, the descriptives and normality of the data were checked, to decide the type of analytical techniques to be used. Then, the means of each variable were compared to examine the developmental profile of the social network characteristics under normal SA circumstances. Next, the friends' characteristics were analyzed to illustrate a profile of the typical friends that the participants made. Lastly, some personal factors which could influence the social network features were examined.

5.1.2.1 Descriptives and normality

5.1.2.1a T1 and T2 data

After screening invalid data, 133 cases were kept for descriptive analysis. At T1, as shown in Table 5.9, the participants on average had 14 L1-speaking friends and 8

English-speaking friends. They had about three English-speaking social circles. They had discussed at least 5 different topics in conversations with English-speaking friends and engaged in 14 English-speaking activities in a month. Respondents on average listed 1.48 English-speaking friends with their biodata and they had nearly two ties among their listed friends. Based on the coding protocol in Table 3.14, their relationship with the listed friend was quite good and their friends' English proficiency was slightly better than theirs. They met their friends offline and talked to them online every week. Histograms showed that none of the variables reached normal distribution, because most network variables were positively skewed and the friend's number variables had a large range, creating big gaps in between.

At T2, the participants on average had 16 L1-speaking friends, 9 English-speaking friends, and nearly three social circles (Table 5.9). They had discussed at least 5 different topics when chatting with English-speaking friends, whereas they participated in 12 activities in English in a month. Respondents listed 1.66 English-speaking friends on average, and their English-speaking network's density was still about two connections. Their relationship with friends was quite good and their friends' English proficiency was mostly better than theirs. They met their friends face-to-face every week and they talked online slightly less often than that. Histograms showed that none of the variables were normally distributed for the same reason as in T1. Thus, all the SASNS variables will use non-parametric tests in further analysis.

Table 5.9 Descriptive statistics of network variables in the SASNS

Variables	T1			T2		
	Mean	SD	N	Mean	SD	N
L1 network size*	13.92	12.98	128	16.40	14.22	144
Network size	7.96	9.78	118	8.89	12.25	132
Network dispersion	3.32	2.92	95	2.65	2.34	114
Multiplexity of topic	5.18	3.41	133	5.06	3.11	142
Multiplexity of activities	14.48	19.92	134	11.89	18.77	140
Network density	1.84	2.387	105	2.15	2.345	125
Friendship intensity	2.95	1.13	102	3.08	.94	123
Friends' proficiency	2.47	.90	102	2.84	.77	123

Frequency of face-to-face contact	3.21	1.02	102	3.13	.96	123
Frequency of online contact	2.90	1.18	102	2.77	1.07	123

Note: * marks L1-speaking item. The rest are English-speaking network features.

In terms of the multiplexity measures, the total frequency of students attending activities in English and the total frequency of their topics of discussion in English in the past month at T1 and T2 are displayed in Figures 5.5a and 5.5b. At T1, sharing meals was the most frequent activity that they did in English, followed by shopping and having coffee. As for topics, the top three most frequently-mentioned topics were schoolwork, hobbies, and home countries. At T2, sharing meals and having coffee were still the most frequent activities that they did in English. Their need for shopping was not as urgent as when they first arrived and they attended more social activities. As for topics, the top three most frequently mentioned topics in English were still schoolwork, home countries, and hobbies. The frequency of discussing gossip increased, suggesting the students were more familiar with each other.

Figure 5.5a SASNS activity frequencies by rank

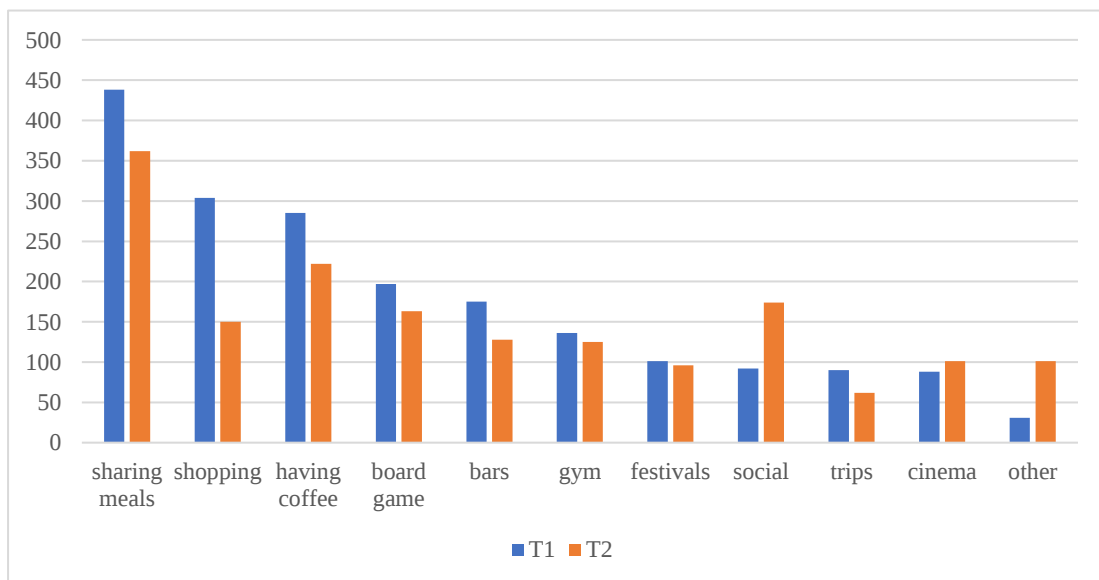
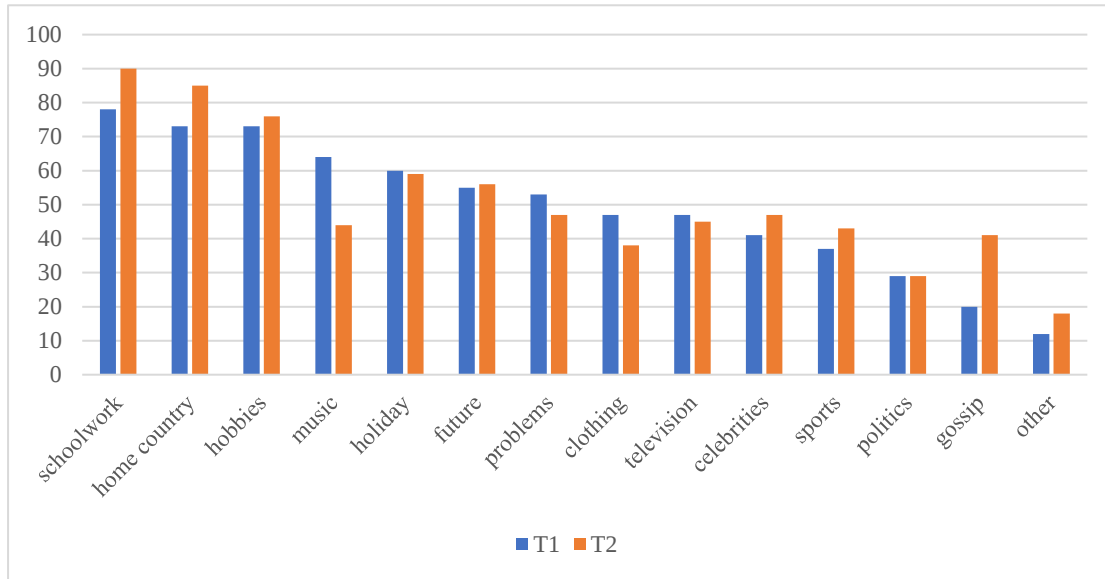


Figure 5.5b SASNS topic frequencies by rank



5.1.2.1b T1 to T2 gains

The SASNS gains were calculated by the T2 value minus the T1 value of the same variable. The descriptive of the gains are presented below (Table 5.10) and there was very little change in each variable from T1 to T2. What is consistent between variables was the large individual differences. The gains variables were symmetrical which implied normal distribution but many were leptokurtic because the change of many participants was close to 0. Therefore, non-parametric tests were suitable for the gain variables. The sample size was small because 68 T1 participants did not fill in the SASNS at T2 and 72 T2 participants did not fill in the T1 SASNS.

Table 5.10 T1 to T2 gains of the SASNS

Variables	N	Mean	SD
Number of L1 friends gain*	64	-.64	9.86
Number of English-speaking friends gain	54	.28	5.14
Network dispersion gain	40	.35	1.39
Multiplexity of topics gain	67	.37	3.29
Multiplexity of activities gain	70	-2.01	16.23
Network density gain	47	.47	2.25
Friendship intensity gain	47	.35	0.93
Friends' proficiency gain	47	.38	0.95
Frequency of face-to-face contact gain	47	-.32	0.91
Frequency of online contact gain	47	.03	1.28

* L1-network item

5.1.2.2 Comparison of means

Non-parametric tests (Wilcoxon-Signed Rank test) were used in the comparison of means of social network variables between T1 and T2. After Bonferroni correction (p threshold = $.05/10 = .005$), none of the variables reached significance (see Appendix I for the results). Because the social integration level of the sample was consistently low during the SA period, the SASNS variables did not capture significant changes in the sample.

5.1.2.3 Social network characteristics

The above analysis showed that the sample had a limited English-speaking social network. In this section, a few characteristics of their English-speaking friends (e.g., cultural backgrounds, means of encounter, friendship intensity, and friends' English proficiency) are analyzed to illustrate a typical profile of the English-speaking friends they listed.

5.1.2.3a Cultural background

The cultural distance between the respondents of SASNS and their friends was coded in two ways, the geographical distance and the cultural distance. In the geographical categorization, the friends' home country was coded into three types: from a country in the same continent as the respondent (except the UK), from a country in a different continent as the respondent, or from the UK. The cultural categorization coded the nationality of their contacts into the same three types based on the circles of English by Kachru (1989), being the Expanding Circle, the Outer Circle, and the Inner Circle (Figure 2.11 in the literature review). Specifically, Hong Kong was coded as the Outer Circle because of its colonial history.

The geographic distance and the cultural distance of the friends that the students listed are presented in Table 5.11a and 5.11b respectively. It showed that students made friends from all over the world with a similar percentage of friends from the same

continents as themselves and friends from other continents, but they did not have many British friends. Regarding the cultural distance from the UK, most friends were from the Expanding Circles of English, with the smallest percentage from the Inner Circle of English. The number of friends from the Outer Circle fell between the Inner Circle and the Expanding Circle. Simple subtraction showed that the number of non-British Inner Circle friends at T1 and T2 were 2 and 10 respectively.

Table 5.11a Geographic distance of listed friends

	T1		T2	
From the UK	8	6%	25	12%
Different continent	54	43%	84	41%
Same continent	64	51%	96	47%

Table 5.11b Cultural distance of listed friends

	T1		T2	
Inner circle	10	8%	35	17%
Outer circle	28	22%	40	19%
Expanding circle	87	70%	133	64%

Note: Inner circle includes the UK

5.1.1.3b Other features

In terms of the first encounter with their friends, the most common way they made friends was through attending classes (Table 5.12). Since the sample was students, it is not surprising that most of their friends were their classmates. The second most typical way was making friends with flatmates, followed by meeting through friends' introduction and meeting through student activities. Other means of first encounters require strong initiatives from the respondents (e.g., online apps, public events), and the chance of making friends in these ways was small for this sample. Based on data of all the friends that the respondents listed at T1 and T2, they preferred to list friends whom they had very good relationships with (Table 5.13), and their listed friends were mainly non-native English speakers with higher English proficiency than the participants (Table 5.14).

Table 5.12 The first encounter of the listed English-speaking friends in the SASNS

Means of first encounter	T1		T2	
We study together	98	65%	132	65%
We live together	23	15%	23	11%
Through student activities	5	3%	12	6%
Through friend's introduction	10	7%	8	4%
We met online	2	1%	1	0%
We met in public events	4	3%	10	5%
We work together	2	1%	5	2%
Other	7	5%	11	5%

Table 5.13 Friendship intensity in the SASNS

Network intensity	T1		T2	
Not so close	19	13%	12	6%
Just friends	33	22%	53	25%
Good	30	20%	43	20%
Very good	70	46%	107	50%

Table 5.14 Friends' English proficiency in the SASNS

English proficiency	T1		T2	
Lower than me	21	14%	21	10%
Similar to me	46	30%	45	22%
Better than me	61	40%	99	49%
Native	23	15%	38	19%

5.1.2.4 Factors influencing social networks

In this section, the same as for the LCQ, several personal factors that could potentially influence the social networks of the participants were examined, to answer RQ2a. The independent factors ranged from nationality, age, English proficiency, motivation, living arrangements, to center locations. Analyzing these factors could also facilitate the construction of the final mixed-effects model. The same as in Chapter Four, Bonferroni correction was applied to multiple correlation analyses (p threshold = .05/10 = .005).

5.1.2.4a Nationality

Firstly, the Mann-Whitney U test was performed on the Chinese and non-Chinese participants. At T1, three variables were significantly different (Tables 5.15a). Chinese students had significantly more L1 friends than non-Chinese participants ($Z = -3.994, p < .001$), whereas non-Chinese participants had significantly more English-speaking friends ($Z = -3.630, p < .001$) and they had significantly more dispersed English-speaking social networks than Chinese participants ($Z = -3.682, p < .001$). At T2, two variables showed significant differences (Table 5.15b). To be specific, Chinese participants had significantly more L1 friends ($Z = -4.488, p < .001$); friends' relative English proficiency was higher for the Chinese participants ($Z = -4.056, p < .001$). Nationality did not have an impact on the network change variables between T1 and T2. Overall, Chinese students had a clear disadvantage in making international friend compared with their non-Chinese counterparts.

Table 5.15a Nationality difference in the T1 SASNS variables

Variables	Mann-Whitney U	Wilcoxon W	Z	p
Number of L1-speaking friends	863.500	1458.500	-3.994	.000
Number of English-speaking friends	859.000	4262.000	-3.630	.000
Network dispersion	434.000	2990.000	-3.682	.000

Table 5.15b Nationality difference in the T2 SASNS variables

Variables	Mann-Whitney U	Wilcoxon W	Z	p
Number of L1-speaking friends	841.000	1337.000	-4.488	.000
Friends' relative English proficiency	698.500	1104.500	-4.056	.000

5.1.2.4b English proficiency

The initial English proficiency (IELTS) and overall English proficiency (VLT) were correlated with SASNS variables at T1, T2, and the gains variables between T1 and T2 in Spearman correlation (Appendix I). At T1, students with higher IELTS scores had slightly higher network density ($\rho = .339, p = .004, N = 71$), while students with higher

VLT scores considered the relative English proficiency of their friends lower ($\rho = -.284$, $p = .004$, $N = 99$). At T2, the VLT scores yielded two significant weak correlations: those with lower VLT scores tended to have more L1-speaking friends ($\rho = -.317$, $p < .001$, $N = 132$), while those with higher VLT scores had a higher L2 network density ($\rho = .323$, $p < .001$, $N = 114$). Overall, more proficient students seemed to have higher quality in their L2 social networks.

5.1.2.4c Motivation

Ideal L2-Self motivation score was correlated to SASNS variables at T1 and T2. Spearman correlation showed that motivation score had a moderate positive relationship with the network dispersion level ($\rho = .353$, $p = .001$) and the face-to-face contact frequency with their English-speaking friends ($\rho = .315$, $p = .002$) at T1. The motivation score did not show a significant correlation with L2 social network variables at T2 (Appendix I). Again, because motivation data was only collected at T1, the lack of correlation with T2 SASNS variables suggested the highly dynamic nature of motivation.

5.1.2.4d Age

Age ($M = 21.07$, $SD = 3.59$) could be a factor that influenced the type of friends that students made. At T1, two pairs of correlations were significant. Firstly, age had a negative correlation with the number of English-speaking friends ($\rho = -.393$, $p < .001$). In other words, the older a participant was, the fewer foreign friends he/she had. Secondly, age had a negative correlation with the dispersion of the social networks. It was found that younger students had a more dispersed social network than older participants ($\rho = -.391$, $p < .001$). At T2, no correlation was found between age and social networks (Appendix I). Overall, younger participants seemed slightly more active in making foreign friends when they first arrived in a new country.

5.1.2.4e Living arrangements

Regarding living arrangements, the Kruskal-Wallis test was employed to examine

whether living by oneself, living with co-national friends, or living with international friends had an impact on the social network features of the participants. At T1 and T2, none of the variables saw a significant difference after Bonferroni correction (see Appendix I). The only variable approaching the significance threshold was L2 network density, with those living alone having lower network density than those living with international friends: $H(2) = 10.532, p = .005$.

5.1.2.4f Center locations

Lastly, the role of center locations in influencing social network variables was explored through the Mann Whitney U test. At T1 and T2, no significant difference was found in the social network variables between center locations (see Appendix I). This result revealed that the social network patterns among the participants at the two centers were similar, probably due to the high proportion of Chinese in the student populations of these two cities.

5.1.2.5 Summary of the SASNS data

The above analysis showed that the participants always had more L1 friends than English-speaking friends under normal SA circumstances and the SASNS did not capture any significant change in their L2 networks throughout the first two terms. The most common activities they did with L2 friends were having meals and having coffee and they talked about schoolwork and their home countries the most often. They went shopping with their L2-speaking friends less at T2 and attended social activities and talked about gossip more at T2. Most of the English-speaking friends the students made were their classmates whom they met weekly face to face. The friends of the respondents usually came from the countries in the Expanding Circle of English but had better English proficiency than the respondents.

The Chinese nationality had a strong relationship with the sizes of English-speaking and L1-speaking social networks and the dispersion of L2 social networks. Non-Chinese students had more English-speaking friends, fewer L1-speaking friends, and

higher L2 network dispersion. Overall English proficiency, motivation, and age had a minor influence on social networks. It was found that more proficient students had more English-speaking friends and higher English-speaking network density. Students with higher motivation had more dispersed L2 networks and met their friends more often. Younger students had more English-speaking friends and more dispersed social networks at T1. The center location and living arrangement did not have a significant impact on the characteristics of the social networks of the participants.

5.1.3 Interview data

Interview data from 29 interviewees were analyzed through thematic analysis to reveal their informal L2 contact patterns under normal SA circumstances. This section is broken down into general language contact and L2 social networks. An interviewee profile with their LCQ scores and interview attendance is shown below (Table 5.16). Their ID consists of three parts: the center location (A/B) + Nationality's initial (Chinese/Ivorian/Egyptian/Korean) + number. Their language contact score was color shaded, with the darker color suggesting higher scores and the lighter color signaling lower scores. The interview attendance at each time point was marked with a tick.

Table 5.16 Interviewee profile by LCP scores

ID	Gender	Nationality	LCP1	LCP2	LCP3	T1	T2	T3
AC1	Male	Chinese	42			✓		
AC2	Female	Chinese	40	43	35	✓	✓	✓
AC3	Female	Chinese	27	34		✓	✓	✓
AC4	Female	Chinese	58	51	35	✓	✓	✓
AC5	Male	Chinese		36		✓	✓	
AC6	Female	Chinese	66			✓		✓
AC7	Male	Chinese	30	40	40	✓	✓	✓
AC8	Male	Chinese	22			✓		
AC9	Female	Chinese	34	42	39	✓	✓	✓
AC10	Female	Chinese	31	31	25	✓	✓	✓
AC11	Female	Chinese	44	49	34	✓	✓	✓
AC12	Male	Chinese	47	47	44	✓	✓	✓
AC13	Female	Chinese	27	34	28	✓	✓	✓
AC14	Female	Chinese	43	55	37	✓	✓	✓
AC15	Male	Chinese	31	37		✓	✓	

AC16	Female	Chinese	44	35	40	✓	✓	✓
AC17	Female	Chinese	50			✓	✓	✓
AC18	Female	Chinese	54	32		✓	✓	✓
BC1	Female	Chinese	37	46	35	✓	✓	✓
BC2	Female	Chinese	41	40		✓	✓	
BC3	Female	Chinese	29	40	36	✓	✓	
BC4	Male	Chinese	53	44		✓	✓	✓
BC5	Male	Chinese	57	65	56		✓	✓
BE1	Female	Egyptian	59	61	48		✓	✓
BC6	Female	Chinese	31	29	32		✓	✓
BC7	Female	Chinese	38	34		✓	✓	✓
BC8	Female	Chinese	62	61	54	✓	✓	✓
BI1	Female	Ivorian	37	48	57	✓	✓	✓
BK1	Male	Korean	31	34		✓	✓	

Note: 1. Yellow shaded students were in China at T3.

2. The second letter: C=China, K=Korea, I= Ivory Coast, E=Egypt

5.1.3.1 Setting the scene

To understand the students' backgrounds better, their reasons for SA were asked as a warm-up question at T1. The varied reasons could be summarized into two: instrumental motivation and intrinsic motivation. For instrumental motivation, some students wanted to avoid the highly competitive College Entrance Exam in China (n=2) or failed the Postgraduate Entrance Exam in China (n=2), so they studied abroad for degree progression. Some wanted an overseas degree to enhance their career prospect (n=3). Only two interviewees said improving English was their main goal of SA (n=2). As for intrinsic motivation, personal interest in the British culture, the hope to broaden their horizon, and the desire to experience a different lifestyle were frequently mentioned by students (n>15). An unusual reason was that Participant AC8 and a few other students studied abroad because he behaved badly in schools in China, so his parents sent him abroad hoping that SA could redeem their children's unsuccessful upbringing.

The reason for most students enrolling in this specific foundation program is that they were not academically prepared for the degree study in university. For example, some students missed the university application deadline when they decided to study abroad

(n=3); some wanted to change their majors but could not get an ideal university offer directly (n=2); some working professionals wanted to take this foundation program as a transition before going back to university (n=2). One participant could not complete her study in the US due to visa rejection. Thus, she came to the UK to continue her study and attended this program for transition. Although a few interviewees (n=11) attended international schools or had lived abroad before, overall, the participants in this foundation program could be described as SA students who needed academic support, compared with those who could get a good university offer directly.

5.1.3.2 Language contact

In this section, above all, an overview of the informal language contact of interviewees reveals the extent of their immersion in the L2-speaking environment. Afterward, the categories of the most frequently-mentioned L2-using activities are presented. Finally, the developmental profile of the language contact in the first six months of their SA sojourn (i.e., between T1 and T2) is shown. All the quotes in English were extracted verbatim and the grammar mistakes that do not impede comprehension were not corrected.

5.1.3.2a Overall language use: $L1 > L2$

In general, students used more L1 than L2 in their spare time at T1. The majority of interviewees (77%, n=20) reported more L1 use than English use, mostly because their friends were compatriots. For example:

In the school, our classmates are all Chinese. Although in the beginning, we said “no Chinese, only English in the class”, people are still speaking Chinese in the class, because they are more comfortable to speak their own language. I haven’t met any foreign students in my class. I even met more foreigners in Beijing than here, so I am like: “Am I really abroad?” (BC8)

This excerpt showcased a typical example that students used their L1 a lot both in class and after class. Two interviewees even reported self-perceived attrition in English

fluency during SA, because they spoke English “*at most a few sentences per day*” (BC4).

Five interviewees (19%) reported more English use than L1 use and one student (4%) reported an equal proportion of L1 and L2 use. Among these five interviewees, Participant BI1 from Ivory Coast had the most immersive SA experience. Her closest friends were all from Africa but did not share the same L1 as her, so they communicated in English all the time. The other three interviewees that used English a lot had international flatmates. Participant AC4 lived by herself and but she tried to spend time with the only non-Chinese classmate in her class. It seems that it required substantial personal efforts for Chinese students to use L2 more than L1, due to a large number of compatriots.

5.1.3.2b Specific L2 use: academic, leisure, and miscellaneous activities

The activities that students used English in were summarized into three categories, being academic activities, leisure activities, and logistic/miscellaneous activities. The main type of out-of-class English contact of them was academic activities, which coincided with the high homework writing frequency in the LCQ results. This category includes both interactive English contact, such as speaking with teammates in group work ($n > 15$), and non-interactive English contact, such as writing emails to teachers and universities (AC17), preparing for the IELTS exam (AC14), reading textbooks or other course materials (AC9), reading the Newsweek and listening to BBC 6-minute English (AC10). Usually, more diligent students had more academic contact. As Participant AC17 commented: “*It’s not necessary but if I read it, I can get higher scores for my courses... so I read them*”. As illustrated below, the characteristics of this type of L2 contact are the utilitarian purpose and the temporal nature:

[I use English most often] for group discussion... In our group there is a Korean student, so we use English with him. It's often before the exam, usually every day in the week before the exam. But after the exam, only once or twice a week. If we meet, it usually lasted for two to three hours... We just compare notes, see whether

we wrote down what the teacher said in class in the same way. [The topic is] limited to academic stuff (AC9).

The participants did these activities for the utilitarian purpose because they were either required by the syllabuses or following the suggestions of their English teacher, not voluntarily for their genuine interest. The topics were usually restricted to academic matters. As for the temporal nature, most of the academic goals that the students prepared for had a definite finishing time (e.g., submitting assignments by the deadline; reading the course textbook before the next class; doing IELTS practice before the IELTS exam), which affected the amount of time they spent on these activities.

The second category of informal L2 use was leisure activities, which included socializing with friends, watching TV series and movies, watching YouTube, listening to music, and so on. A few findings emerged from this strand of data. Firstly, the frequency of going out with English-speaking friends showed a polarizing trend. Some participants had never hung out with friends speaking English (n=7). As Participant AC3 explained: *“Even if we have one or two foreigner classmates, I usually go out with the Chinese one because I feel very comfortable and warm with them”*. In stark contrast, some went out with English-speaking friends every week (n=3), such as the Ivorian girl who said: *“I use only English all the time. We hang out. We go to the restaurants, we go shopping, we have lunch together 3 or 4 times in the week. Basically, we spend all the time together. It's a group, a circle” (BII)*. Secondly, the proliferation of video websites has a profound impact on the lifestyle of young people, so they can easily access the contents they like. They mentioned watching movies online (n=12) and watching YouTube often (n=4). The videos they watched ranged from car vlogs, beauty, and make-up vlogs, to inspiring TED talks, depending on students' interests. Thirdly, the frequency of leisure activities is consistent and habitual. In other words, students either spent time on certain hobbies regularly or hardly ever did it. For example, Participant AC15 shared: *“I usually play [games] for one or two hours before sleep; on the weekend I probably play a little bit more, about 10 hours in total”*. Although

only four interviewees mentioned playing computer games in English, they did it regularly as part of their daily routines. The patterns of L2 contact were triangulated by their answers in the IELTS speaking test, which highlighted the prioritization of academic tasks and the high frequency of L2 receptive activities.

Lastly, miscellaneous activities were what the students could not avoid and sometimes felt challenging and overwhelmed about in a new living environment, such as opening a bank account (n=3), registering with the GP (n=2), solving problems with the receptionist at their accommodation (n=6), which required speaking English with strangers. For example:

It's my first time come to the UK. There are many problems I never encountered before. For example, I can't use bank card for online shopping, so I have to go to the bank to find out what's the problem with it. I think it is very difficult for me because there are many words I don't know (AC5).

It revealed that students were often daunted by the unfamiliar technical terms used in these commitments. Other logistics activities such as ordering food and shopping were frequent and relatively easy for the participants. But sometimes they also encountered small issues, which left them a deep impression. For instance, Participant AC14 recalled: “When I first arrived in the UK, I went to a shop to buy plates. But I said ‘plant’, so they can’t understand. I used my body language and then they finally understand it’s ‘plate’”. Embarrassing experiences like this were not unusual in the interview, which explained why many students lack the confidence to speak English in the first term, be it with foreign classmates or strangers.

5.1.3.2c The developmental profile of students’ L2 contact: mostly stable

The developmental profile of informal language contact is what the present study is primarily interested in. Consistent with the quantitative data, among the 23 interviewees who took part in both rounds, most interviewees maintained a similar level of L2 contact from T1 to T2, but they had more study time and less social time/free time

(n=13). In British universities, the second term of an academic year is usually the most intensive. A few students with established English-speaking networks at T1 expanded or deepened their friendship at T2, while many others just gave up on making foreign friends and focused on studying instead. The data will be presented in the same categories as above, namely, academic, leisure, and miscellaneous/logistic activities.

First, study-wise, many interviewees reported more L2 use due to course arrangement (n=7). This was not only evidenced by information like *“last term we had one assignment and this term we have three assignments”* (BC8) but also by remarks such as *“busy/busier”* (n=5) and *“more tired”* (AC16) due to the coursework. Their language contact increased because they took their study seriously and used English more for academic purposes. Interestingly, two interviewees at Center B said they paid less attention to the English class because they concentrated on the subject classes. Participants in Center A needed to obtain IELTS scores for university applications, so they spent considerable time preparing for the exam.

Seven interviewees at Center A used English more than T1 because of IELTS preparation. For example, Participant AC3 was highly motivated and paid for expensive online one-on-one IELTS classes. As she reported: *“I practice IELTS more. I use 3-4 hours every day... I get better results in IELTS. I very enjoy the sense of achievement”*. This excerpt showed that she had a pragmatic orientation with regard to English learning. It should be noted that although they spent more time preparing for IELTS, they admitted using Chinese as much as before (n=2). Without the compulsory IELTS requirement, four interviewees at Center B maintained the same intensity of L2 contact as T1. These examples showed that external academic requirements could increase the non-interactive L2 use, but not communicative L2 use of this sample.

Psychologically, after one term's settling-in, students were more adjusted to the SA life and Western teaching styles. While many participants still felt stressed about their studies or university application and kept trying their best (n=11), few low-proficiency interviewees just took it easy and only wanted to pass with minimal grades (n=2).

Participant AC8 openly shared: *“My biggest change is that I care about the exam less... I look for the most cost-effective way to pass”*. The learners with this attitude did not increase their study time. Participant AC16 had an in-between attitude with a wish to work hard, but did not put it into action. She reflected: *“I am quite procrastinated...I want to improve my English so I downloaded many apps. But I didn’t use them. I am in two minds whether to start this”*. This excerpt showed the mental swing of a few learners on whether to put their full efforts into study.

Second, the amount of English use in leisure activities was stable from T1 to T2, both in interactive contact (e.g., hanging out with friends) and receptive contact (e.g., watching movies). On one hand, their interactive contact and social network were mostly unchanged. Most interviewees did not form close English-speaking friendships at T1 and this phenomenon persisted till T2. For instance, Participant AC3 only had Chinese friends. She said: *“I don’t use English more in my daily life. I think I have more Chinese friends this term. To be honest, it’s few opportunities for me to make friends because we are just in our existing social circle. We study together and play together”*. The few interviewees who established English-speaking friends’ circles at T1 maintained close contact with the same friends. For example, Participant BI1 said: *“It’s the same. I use English every day, because of my class, my friends”*. On the other hand, their receptive leisure contact was also stable. For instance, Participant AC5 shared that he did not change his hobby of playing computer games from T1. These hobbies were not formed because of studying abroad. Rather, they were habits that the students formed over the years, which they always enjoyed.

Finally, students generally got better at handling miscellaneous activities and they were more confident in speaking with service personnel. As Participant AC9 reflected, her biggest change at T2 was *“not scared to speak English with other strangers”*. *“When it’s necessary, don’t be afraid. Just speak.”*, which was considered *“a big step”* by her. With more details, Participant AC15 gave some examples:

When I first arrived, I tried to avoid speaking English. For example, when shopping in the supermarket, I look for things by myself. When I order in restaurants, I first decide what I want and then tell the waiter directly. This term I am able to speak more...Now if I want to look for something, I will ask the shop assistant. When I order, I will ask the waiter how is the dish. This term I started to order food delivery on the phone; I reported some repairs to the reception; I posted my ID card back to China and my family got it in the end.

This progress is an encouraging finding, which showed that despite limited L2-speaking friends circles, some students improved English communicative competence through running errands and solving daily problems.

Overall, most participants still reported more L1 use than L2 use at T2. The spoken English contact of the sample was still polarized, with only a handful of participants having close English-speaking friends (see Section 5.1.3.3) and most students were stuck with strong L1-speaking social circles. Because much of the increased L2 contact was academic-related, students' other casual or logistics English use largely remained unchanged or decreased. Lastly, students had more confidence to handle miscellaneous issues in English.

5.1.3.3 Social networks

L2 social network is a key indicator of the quality and frequency of L2 speaking contact of SA participants. In this section, the overall friendship patterns of participants will be outlined. Since L1-speaking friends are not the focus of the present study, the details of the international friends and British friends, and the developmental profile of L2 networks from T1 and T2 will be presented, with two cases as vignettes.

5.1.3.3a General social networks: L1 friends > L2 friends

Overall, most of the interviewees (n=22) reported more and closer L1-speaking friends than English-speaking friends, while only few interviewees (n=4) had more English-

speaking friends than L1-speaking friends, which was in line with the SASNS results. Typically, their co-national, international, and host-national friends were situated in the inner, middle, and outer circles of their friendzone. As Participant AC17 put it:

The most close friends are Chinese. In the second round of my relationship, that is foreigner. In my classmate, there is Russian, Malaysian... We talk to each other in English. Because of some cultural difference, we cannot build close connection with each other to some extent.

This excerpt revealed that even if they had English-speaking connections, the relationship was weak. In a similar vein, Participant AC5 recognized there was “*an invisible circle that limited your communication*” and he had just talked with several foreigners so far: “*the students, the receptionist, the bank cashier, the retailers, the teachers*”. Most students did not have access to local friends. Further, some interviewees only had co-national friends but no international or local friends.

The phenomenon of co-nationals clustering applied to all nationalities in the program. For example, students from the same country tended to sit together in class. As Participant BC3 observed: “*Even if I want to make Asian friend, you will notice Koreans only hang out with Korean, Japanese only hang out with Japanese*”. Interestingly, she speculated that Chinese students had low approachability to non-Chinese students. “*Because I am already surrounded by Chinese friends, foreign friends may feel your social circles is full already. They don’t wanna join...*” Thus, the close-knit L1-speaking circles might potentially trade off the possibility of L2 friendship of the Chinese students.

Reasons given for mainly having L1-speaking friends by the interviewees were both sentimental and pragmatic: a sense of security; getting help quickly; avoiding misunderstanding; feeling more relaxed and comfortable (e.g., similar living standard); curing homesickness and loneliness; talking about news in their home country; maintaining cultural identity; establishing connections after they return to their home countries. Although some Chinese interviewees (n=2) complained that the relationships

between their Chinese classmates were too complicated, most interviewees deemed the benefits of L1-speaking friendship important for them.

5.1.3.2b International friends

The interviews provided rich details which aligned with the friends' portraits in the SASNS. Generally speaking, the interviewees usually made friends with foreign classmates from the start of the term (e.g., they sat next to each other at the Welcome Week events; they were assigned to the same group in a course, n=7). As their friendship was superficial (i.e., just greeting each other when they met) and they did not know each other very well, many interviewees did not have the contact details of their foreign classmates/flatmates. They sometimes had small talks before class and after class on academic topics, such as "*assignment submission, the scores and revision*" (BC7), and "*the foundation program and education in general*" (BE1), but they did not hang out together. They rarely made plans with international friends outside the campus, except for doing group work. Going to pubs (n=4) and sharing meals (n=5) were the most common leisure activities that they did following the group work.

On the contrary, it is less common for students to become friends with their international flatmates. Although some flatmates shared a kitchen, they did not meet often or talk much due to different timetables. Three interviewees even reported bad relationships with their foreign flatmates because of different living habits. For instance, two interviewees shared unpleasant incidents with their American flatmates because of kitchen cleaning and noise disturbance respectively. Moreover, some students did not socialize with their flatmates because they were tired after a long day's study and only wanted to relax in their own rooms (AC4).

The reasons behind the difficulty of making international friends are complicated. Above all, limited English proficiency affected smooth communication and some interviewees had difficulty in understanding non-native English accents (n=5). Further, some interviewees underscored the cultural differences, which meant that they often ran out of topics to chat about and did not have many mutual interests (n=4). What

stood out in the data was how Chinese interviewees perceived the pub culture in the UK. For those with more international friends, “*go to the bar to talk is the most common*” (BC8), and “*drinking is a must*” in socializing (AC12). However, Participant BC7 disagreed: “*Some people say you can go to bars [to make friends], I think it's too noisy to talk. Do I need to shout? So I give up going to pubs. Now I just let it be*”. In China, it is uncommon for students to go to bars. This contrast suggested the importance of intercultural awareness. Furthermore, another reason might be the lack of time and motivation due to academic work. As Participant AC4 commented:

I don't have that energy to go out and meet new people. I think it's because I have stress about applying schools so I don't want to go out. Even though I won't do anything at home, I will just feel guilty if I go out and have fun.

This example underscored that degree-seeking international students were constantly trapped in a dilemma of spending time studying or having fun. Many chose to study because of the high stakes of their university applications.

In the interview, there were a few successful exceptions of international friendships (n=4), almost exclusively between friends from a similar cultural background with shared values (e.g., Chinese-Korean, Ivorian-Nigerian). The commonality of these friendships was that they met regularly because of university or work, so they had lots of opportunities to get to know each other. Besides, these participants tended to be outgoing, adventurous, easy-going, and open-minded. The following vignette illustrates a cross-cultural friendship example between China and Korea in detail:

Vignette 1

Participant AC12 was a 29-year-old Chinese male student with IELTS 6.0 at the start of the foundation program. He had worked in a multi-national firm and some NGOs in China. He was gay and he paid attention to LGBT activism. His hobby was travelling and he had been to over 20 countries before he studied abroad, while some of his Chinese classmates had never travelled abroad. According to other interviewees, he was warm-hearted and had strong empathy. He had many Chinese friends and international friends because of his personal charm. At T1, he was close to a French classmate but his best friend was Chinese. At T2, a Korean girl became his best friend. He said: *“I knew her since last term, but we were not close. Since this term we become super close. One reason is that we attend almost the same classes. Another reason is that she is of similar age as me. She is also a sentimental person. We have a lot to talk about”*. This example reveals that being a few years older than most classmates, AC12 naturally gravitated toward people who shared most commonalities with him and could provide equivalent emotional value.

Lastly, the following benefits of having English-speaking friends were recognized by the participants, such as broadening the horizon, improving English, and learning different ways of thinking. For example, Participant BC3 thought that India was *“underdeveloped”* and *“does not treat women equally”* but she learned from her Indian friend that India had some *“cutting-edge research in the world, such as IT and pharmaceuticals”*, which changed her stereotypes. In university applications, Participant AC14 noticed that Chinese students were obsessed with university ranking and the judgment from employers, while foreign classmates were *“not as utilitarian as us”*. *“When they choose the university, they will think whether they like the major and whether they like the city”*. Therefore, she felt less stressed about her application.

5.1.3.2c British friends

In the interview data, it was rare that the students could make British friends in this foundation program. At T1, the foreigner that Participant BC7 spoke English most often with was the janitor that cleaned her dormitory every week. Some interviewees (n=3) believed that joining university societies would be a good way of making British friends, but they were either not interested (BC4) or too busy for it (BE1). Since they did not have British classmates, their teachers were their closest local connections in the program (n=5). BC7 tried to practice speaking by actively answering teachers' questions in class and AC17 often discussed "*current issues or domestic issues*" with teachers after class. Participant BC7 explained her rationale behind this:

I think this way I don't need to waste time looking for foreign friends. The teacher will always be there. Every week you have a fixed number of hours of class, if you seize this opportunity, there is no excuse that you cannot improve your speaking.

Outside the program, they only had short and brief contact with the locals. For example, some girls were asked for numbers on the street (n=2), while some students used dating apps (n=2), such as Hinge, Tinder, and Grinder. However, due to the cultural difference, students did not report success in establishing enduring friendships in these two ways.

The romance seemed to be the only way for these students to form meaningful deep relationships with local people (n=2). Based on non-exhaustive information, two Chinese interviewees had British boyfriends and one Chinese participant (not among the interviewees) had a British girlfriend. Interviewees with local boyfriends both had some cross-cultural experiences. Participant BC8 met her English boyfriend when she worked with foreigners in China. After she decided to study abroad, her boyfriend came back to the UK with her. Although they broke up in the first term, they were still good friends and shared a house. Participant AC2 came to the UK on a working holiday visa and worked in London for one year before enrolling in this program. She had both British and international flatmates and she often socialized with local and international friends. She met her British boyfriend at a Taiwanese event in London. It seems that both parties in the interracial relationships usually have some cross-cultural awareness.

5.1.3.2d Developmental profile of social networks: stable pattern

In this section, the developmental trajectory of L2 social networks will be presented in three categories, being stabilization, expansion, and shrinkage. Firstly, in line with the quantitative data, the English-speaking social networks of most interviewees were stable from T1 to T2. On one hand, the interviewees who had strong English-speaking social networks (n=3) continued to speak English a lot with their friends and there was no sign of reverting to the “L1-speaking bubbles”. For instance, Participant BI1 used English with her African clique every day, saying that *“I didn’t have new friends. I still mostly hang out with the first term friends”*. Participants AC2 and AC12 also maintained a primarily L2-speaking friends circle and deepened their relationship with one close contact (friend/boyfriend).

On the other hand, stabilization could also indicate a continuation of weak or small L2 social network, marked by stagnation of L2 friendship intensity (n=20). Those who did not have (close) English-speaking friends at T1 did not make much progress and just let it be. For example, the L2 network of Participant BK1 at T2 *“just stays as usual”*. This was echoed by Participants AC9 and AC17 who reflected that they *“meet less people than last term”* because they *“spend more time alone in studying”*. For many foundation students, English-speaking friends was a peripheral part of their life. Therefore, many did not put much effort into maintaining new friendships and meeting up regularly, unless they really got along well. For example, Participant BC3 met an Indian student from the same university in a bank. She shared: *“We just chatted, but never go out... He has invited me to their Diwali festival. I don’t understand their culture, so I didn’t go”*. This example illustrated how this potential friendship withered without mutual efforts. Similarly, Participant AC16 made a Filipino friend from her class at T1 and they hung out a few times. At T2, their relationship did not enhance because her Filipino friend’s mum came to the UK and she did not have time to meet up. Because international friends were newly-made, their priority was often lower than

their family members or old friends. Participant AC4 also experienced stagnation in the L2 social network, as illustrated below:

Vignette 2

AC4 is a Chinese female student in her mid-20s. She got her bachelor degree in the US and worked in China for 1.5 years. She wanted to change a career, so she came to the UK for a master degree. With prior SA experience, she started the program with strong motivation to make good use of her SA time: *“I know the reason why we study abroad. I don’t wanna stay in the UK and learn different accents of Chinese...I want to isolate from my Chinese speaking environment”*. She explicitly stated improving English as a goal of SA, which directed her socialization preferences. From T1 to T2, she only had one L2-speaking friend (a Malaysian classmate) and her main social clique was Chinese. At T1, she often hung out with this Malaysian girl who studied in Australia before: *“We go to China Town quite often. We watch movie together... We like to hang out after class. We are both girls, we both studied abroad before. We have many things in common, and we both like handsome guys.”*

At T2, they still maintained a close relationship, *“because I don’t have a boyfriend and she is in a long-distance relationship...We know each other, so we talk deeper. Sometimes we be the cheerleader for each other. She would share with me her love story and her struggle”*. However, AC4 was worried about her English fluency attrition. Because most of her classmates were Chinese, she spoke Chinese all the time. Despite strong willingness to practice English, she did not manage to expand her L2 network due to busy study and limited access to local people: *“I think I am the only person in my class that like to speak English. Most of my classmates don’t have the confidence to speak... and I don’t have enough friends surrounding me to speak English with, so I am not satisfied at all.”* In order not to waste her time and money in the UK, she started to listen to English broadcast every day to increase her exposure to English.

As for those with expanded L2-speaking social networks (n=2), the expansion only happened on a limited scale (e.g., two new friends) and largely depended on whether they luckily met “the nice people”. For example, Participant AC11 had a better experience in her group work at T2 than at T1. In the first term, her teammates often delayed in finishing tasks and her experience was unpleasant. In the second term, she developed a genuine friendship with her Italian and Arabic teammates in the new group, as illustrated below:

This term every group work was organized in the library and everyone involved in it... Our group contains 6 people. Including me there are only two Chinese in this group...The foreign friends I met are very nice. They can explain it to you, so I can get it. I can know others' joking... Because of the group work, we discuss at least once a week, and we go out maybe twice a week.

In her case, the relationship with these two new teammates evolved from classmates to good friends with emotional bonds. Both sides showed nice personalities such as senses of responsibility and humor, which were attractive traits that made the friendship last. In the case of Participant BC7, her relationship with her two colleagues was formed since she started working in a Chinese supermarket and her colleagues (a Hong Kongese and a British-born-Chinese) wanted to help her improve her English. Again, a mutual goal to work on and a nice personality from the more fluent speakers played a role in a strengthened L2 speaking network.

In contrast, one interviewee reported a sharp decrease in L2 socializing time. Participant AC5 was very sociable at T1 and went to different bars with his international friends. However, his academic performance was poor and he was rejected by a university in his masters' application. As a result, his daily routine completely changed at T2, in fear of not getting any offer, as illustrated by the following excerpt:

I think the biggest change is that I put attention to my study rather than go outside... I reduced the time outside to have fun. I didn't care too much about study last term. This term I've started to begin, I feel very hard and exhausted.

This example showed that foundation students' academic pressure directly affected their social life and academic success seemed to outweigh the importance of L2 socialization.

5.1.4 Summary of pre-COVID informal language contact

The above quantitative and qualitative data illustrated a comprehensive picture of the L2 contact pattern and developmental profile of international foundation students for the first six months of their foundation year. In brief, they generally used more L1 than L2 in their daily life and used English receptively more than productively. They had more L1 friends than L2 friends and their L2 social networks were generally small and delicate. They had some international friends who were also students in the university, while they barely had any British friends. Their language contact differed significantly by nationality, with Chinese students using English less and having fewer English-speaking friends. The only feasible way for them to make British friends was with their teachers or through a romantic relationship.

In terms of developmental profile, most learners had stable L2 using frequency and their L2-speaking social networks were mostly unchanged. Interview revealed that their study time increased at T2 at the cost of decreased social time. Students reported more confidence in accomplishing miscellaneous or logistics tasks by speaking English with strangers. Only two interviewees had strengthened L2-speaking networks, which could be attributed to the positive and kind personal traits of both sides and the regular opportunities to meet in class or at work. One interviewee experienced shrinkage in his L2-speaking network because of academic pressure. Several linguistic and cultural hurdles remained for this learner cohort to establish cross-cultural friendships.

5.2 POST-COVID LANGUAGE CONTACT PATTERNS

The national lockdown in the UK in Spring 2020 happened after T2 and continued throughout T3 data collection. This section of the chapter is devoted to answering RQ2b:

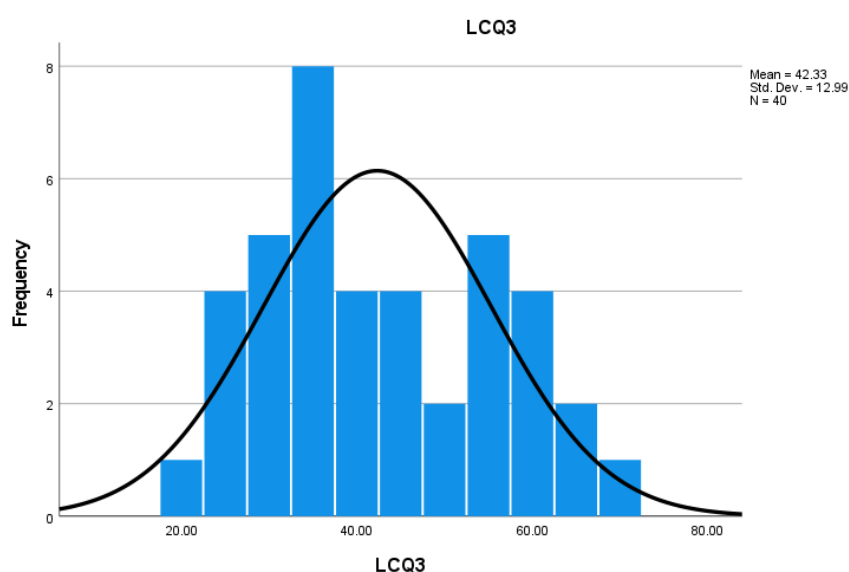
the influence of the pandemic on the SA experience and informal L2 contact of the sample. During the lockdown, all the classes were taught online. Four interviewees flew back to their home countries while the rest stayed in the UK in isolation. These factors made the T3 findings unapplicable to normal SA circumstances, but they provided valuable insight into how the SA experience was fundamentally changed by the external environment (i.e., the pandemic). The data are presented in the same order as above, from the LCQ, the SASNS, to interview data.

5.2.1 The Language Contact Questionnaire

5.2.1.1 Time 3 data

Due to the COVID-19's impact, only 40 T3 LCQ responses were collected. The overall LCQ scores had a mean of 42.33 (SD=12.99). The maximum score was 68, whereas the minimum score was 20. The distribution of normality was checked in the ways mentioned in Section 5.1.1.1. The histogram was visually skewed to the lower end (Figure 5.6), but the extreme cases did not have an impact on the mean score. Therefore, this variable was not transformed.

Figure 5.6 Histogram of T3 LCQ scores



In terms of individual LCQ items, the mean score of all L2 contact was 2.86 out of a possible maximum of 5. The items with the highest frequency were writing homework in English, using social media in L1, and speaking with L1 friends. The participants spent little time speaking with British friends in English, reading books, newspapers, and magazines in English, listening to radio or podcasts in English, and playing computer games in English (Table 5.17).

Table 5.17 Descriptive statistics of the T3 LCQ breakdown

English-contact items	Mean	SD	Min	Max
Writing homework	4.15	0.98	2	5
Using social media in L1*	3.98	1.21	1	5
Speaking with L1 friends*	3.88	1.36	0	5
Listening to songs	3.68	1.07	2	5
Watching movies	3.33	1.27	1	5
Reading timetables, menus, maps, announcements, etc.	3.28	1.24	1	5
Writing Emails	3.28	1.24	1	5
Browsing websites	3.18	1.28	1	5
Using social media	3.13	1.47	1	5
Watching YouTube or Netflix	2.93	1.37	1	5
Speaking with service personnel	2.85	1.49	0	5
Speaking with international friends	2.83	1.71	0	5
Watching TV	2.68	1.64	0	5
Listening to people's conversation	2.55	1.20	1	5
Speaking with British friends	2.38	1.85	0	5
Reading books, newspapers or magazines	2.13	1.24	0	5
Listening to radio or podcast	1.73	1.30	0	5
Playing computer or mobile games	1.58	1.69	0	5

*L1-contact items

5.2.1.2 Comparison of means

A one-way ANOVA was used to compare the overall and item-level language contact difference between T1 (8 weeks after arrival), T2 (20 weeks after arrival), and T3 (30 weeks after arrival). There was a significant effect for time, Wilks' Lambda = .690, $[F(2, 24) = 5.391, p = .012, N = 26]$, with partial η^2 being .31, indicating a large effect size. Post-hoc comparisons indicated that there was a significant difference ($p = .024$)

between the overall LCQ scores at T2 (M=45.46, SD=11.31) and T3 (M=41.00, SD=11.98), with a mean decrease of 4.46. Overall LCQ scores at T1 (M=42.69, SD=11.44) did not have a significant difference from the LCQ scores at T2 or T3.

When it comes to individual item comparison, there were two items of significant differences in repeated-measures ANOVA (Table 5.18), with two increased items and three decreased items. On one hand, the frequency of using social media in L1 saw a significant effect of time (Wilks' Λ =.527). Post-hoc comparisons indicated a significant rise at T2 (p =.002) and T3 (p <.001), compared with T1. On the other hand, the frequency of speaking English with service personnel saw a significant effect for time (Wilks' Λ =.586). The frequency at T3 was significantly lower than that at T1 (p =.002) and T2 (p =.003) respectively. The lower frequencies of speaking English with British friends and listening to people's conversations in English and the higher frequency of watching movies were not significant after Bonferroni correction (Appendix I). Overall, these results suggested lower L2 interactive engagement and higher L2 receptive contact and L1 engagement at T3.

Table 5.18 One-way repeated-measures ANOVA of LCQ items across three time points

N=26	T1		T2		T3		ANOVA			
Variable	M	SD	M	SD	M	SD	df	F	p	Partial η^2
Using L1 social media	2.65	1.57	3.73	1.43	4.04	1.22	2,24	10.759	.000	.473
Speaking with Service personnel	4.08	1.13	3.81	1.23	2.73	1.51	2,24	8.47	.002	.414

5.2.2 The Study Abroad Social Network Questionnaire

5.2.2.1 Time 3 data

At T3, despite the impact of the pandemic, a small number of participants in the UK (N=22) still completed the SASNS online. Table 5.19 showed that at this point, the

respondents on average had 10 L1-speaking friends and 7 English-speaking friends, with two social groups (network dispersion). On average they listed 1.90 friends and the network density had over two connections in the L2 ego-net. Their relationships with friends were quite close and their friends' English proficiency was still better than theirs. The frequency of face-to-face contact was less than once per week, while their frequency of online contact was slightly higher. Histograms showed that none of the variables were normally distributed and were heavily skewed, which was suitable for non-parametric analysis.

Table 5.19 Descriptive statistics of T3 SASNS

Variables	N	Mean	SD	Min	Max
Number of L1 friends	22	9.68	8.82	1	40
Number of English-speaking friends	21	6.71	6.40	1	25
Network dispersion	19	2.21	1.13	1	4
Multiplexity of topics	21	5.57	3.60	0	14
Multiplexity of activities	22	8.23	21.63	0	95
Network density	21	2.43	2.20	1	9
Friendship intensity	21	2.96	1.08	1	4
Friends' proficiency	21	2.76	0.68	1	4
Frequency of face-to-face contact	21	2.44	1.01	1	4
Frequency of online contact	21	2.66	1.01	1	4

Regards multiplexity measures, the most frequent activity at T3 became sharing meals and shopping, which was feasible for those living in the same building and abiding by government regulations (Figure 5.7a). During the lockdown, the frequencies of going to the gym, bars, and going for trips were the lowest, because these places were not open or some public transportation was suspended. The most frequent topic for discussion was schoolwork, closely followed by future plans, hobbies, and problems that they encountered (Figure 5.7b).

Figure 5.7a T2 SASNS activity frequencies by rank

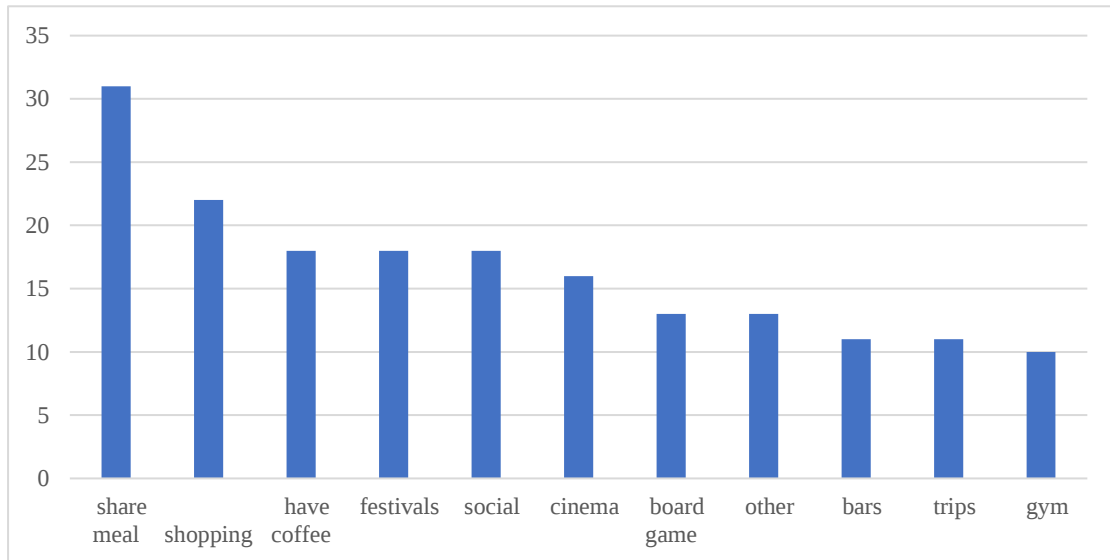
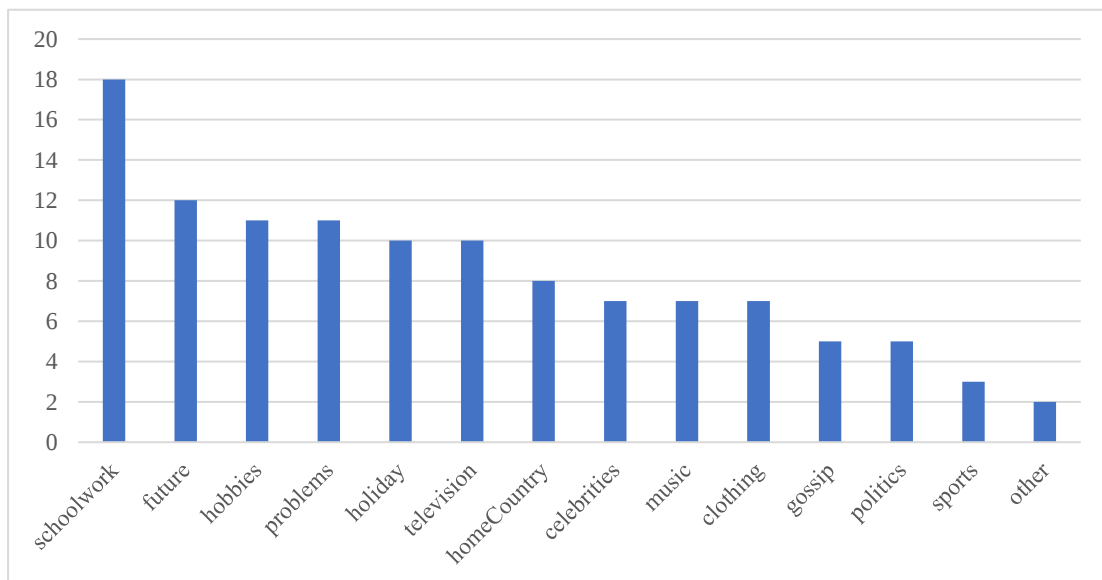


Figure 5.7b T3 SASNS topic frequencies by rank



5.2.2.2 Comparison of means

Non-parametric tests (Friedman's test) were used in the comparison of means of social network variables between three time points. After Bonferroni correction (p threshold = $.05/10 = .005$), none of the variables reached significance (see Appendix I for details). As the social integration level of the sample was consistently low during the

SA period and the sample that completed the SASNS three times was too small (N=13), the SASNS variables did not capture any significant change.

5.2.3 Interview data

22 students' interview data at T3 and some related interview data at T2 were analyzed in this section, to examine how the COVID-19 pandemic affected the SA experience and informal language contact of the UK-based students. The SA experience during the national lockdown was also explored because it sheds light on the profound influence of COVID-19 on individual SA participants' attitudes toward cultural differences and the host country, which indirectly affected their motivation and efforts to integrate into the local community.

5.2.3.1 Informal language contact

The lockdown made the students feel frustrated that part of their SA journey was wasted/ruined. Alongside many restrictions on daily activities, the side effect of the lockdown was the deprivation of interactive English contact. They regretted that they spent money, suffered from fear and loneliness, and got nothing in return. For instance:

The biggest difference of study abroad and in China is that you have more chances to speak English. No matter you want to practice spoken English or not, you practice it unintentionally. But due to the quarantine, although I am physically in London, the people that I contact are in China. I don't have many chances to practice spoken English, so my speaking is worse (AC6).

We have been stuck at home for over a month. We can't go outside. We can't have normal social life. We can't talk to people. We can't even speak Chinese to many people so it's really annoying. I think lots of people are worried about this condition because they don't have a chance to speak out. Study abroad is pointless now (BC8).

As demonstrated in these excerpts, what disappointed the students most was that SA could not offer what they wanted to obtain by residing in a foreign country, such as experiencing the local lifestyle and socializing with people. Language-wise, overall, interviewees reported increased contact with close friends and families in their L1 and less use of English, especially speaking and listening. The data will be presented in the same categories as Section 5.1.3.2, being academic activities, leisure activities, and miscellaneous/logistics activities respectively.

In terms of academic activities, online classes were the main English-use activity for every interviewee (n=22), but the effect was compromised and the quantity and quality of L2 contact decreased. The biggest change was that students spoke English less than before. Between classes, students could not engage in small talks with classmates anymore. During class, speaking was still rare. In the past, students could *“have eye contact with the teacher”* to answer questions (BC7), whereas *“in online class no one turns on the camera”* (AC12). Many students did not want to put themselves under the spotlight, so very few students volunteered to answer teachers’ questions. As a result, some teachers decided to *“allocate the speaking opportunities equally, so they call each student on average once or twice”* (BC7), which made the class atmosphere less active. The second issue related to online classes was the difficulty to concentrate. Without interaction with teachers, students easily zoned out or got distracted in their dorm. Participant AC17 described in a simile: *“It’s similar to watching the YouTube videos without the subtitles”*. Overall, online classes were a challenge to their self-discipline. Some interviewees could not resist using their phones during the class to chat with classmates, which also impaired the quality of this language contact.

Other academic activities that students always mentioned were doing assignments (including searching for literature online and watching YouTube tutorials) and preparing for exams. Participant AC12 shared: *“I am writing essays, so I force myself to read in English. I am also preparing for IELTS. I am not hardworking, but these tasks push you to spend time on English every day”*. These compulsory tasks pushed students to receive reading, writing, and listening L2 contact. The frequency of self-

studying was largely affected by learners' progress in university applications. Some interviewees who already passed IELTS or received an offer from their dream universities did not care about study. For example, Participant AC3 got the offer from UCL and she confessed: *"I finally relax. I completely stopped practicing English since I got the offer because I don't need to take IELTS anymore"*. Those who had not met the language requirement yet studied hard. For example, Participant AC16 voluntarily practiced speaking with her English teacher in order to pass an internal English exam, while at T2 she was not very diligent. It shows that throughout the academic year, the ultimate goal of university application was the main driving force for Chinese students to use English out of class.

Regarding leisure activity, the two biggest changes were increased frequency in browsing news websites and a sharp decrease in socializing time. First, some interviewees (n=5) mentioned they browsed news websites (e.g., BBC, Google, Bloomberg) regularly because they *"really care about the situation of the coronavirus"* (AC17). During the lockdown, students wanted to know the latest official updates on the infection statistics and government policy from the media. Second, face-to-face human interaction reduced dramatically as students could only talk with those who lived in the same household, while online contact with close friends and families increased (see section 5.2.3.2 for more detail).

In terms of non-interactive recreational activity, students had much more time to watch TV series (Netflix), YouTube, movies, listen to podcasts, and play games during the lockdown. However, "binge-watching" sometimes got the participants bored (n=3), as Participant BE1 reported: *"In the last month I watched [Netflix] daily. Now I feel there is no more exciting movie or show to watch, so I am waiting. I have lost interest"*. Similarly, Participant BC5 watched games live streaming online every day. After one month, the lockdown was more than he could bear and he decided *"to go home [return to China] ASAP"*. It seems that these recreational receptive L2 activities did relax the students during the lockdown as pastimes, but because they did not have other options, whether they enjoyed it as much as pre-COVID time was uncertain.

The last type of informal language contact was miscellaneous/logistic activities, whose frequency diminished greatly. Students could not go shopping or have lunch or dinner in restaurants, so they reported having no chance to speak English in miscellaneous activities. When they received food delivery, they tried to speak less with the service personnel to reduce the chance of virus transmission. As mentioned in Section 5.1.3.2, a big improvement of the participants during SA was their ability/confidence to speak English in logistic activities. Unfortunately, speaking development in this regard was interrupted by the lockdown and students felt a big loss.

5.2.3.2 Social network

The social networks of SA participants witnessed evident changes due to the pandemic. Overall, face-to-face interaction with people reduced sharply, except for those who luckily lived with their close friend/partner. The trend of online interpersonal contact was a rise in L1 use and a decrease in L2 use. As Participant BC6 explained: “*when people don’t meet each other, we just choose who we want to talk with*”. Emotional bonds and physical distance were the two main factors influencing who the students contacted. In other words, students were more likely to maintain contact with people whom they had an intimate relationship with AND/OR shared a similar living environment with (Table 5.20).

Table 5.20 Factors that influence contact frequency in the lockdown

Contact frequency	Close relationship	Weak relationship
Short distance	Type 1↑	Type 2 Same
Long distance	Type 3↑	Type 4↓

Many interviewees (n=6) increased L1 online contact with someone that they had a good relationship with and were still physically in the UK (Type 1). For example:

I contact BC1 most often. Because we live in the same city and we share the same time zone, it’s very convenient for us to communicate. I think because we also share the same class, I know when she is free. I will talk with her when she is free (BC6).

Both living in the UK, Participants BC1 and BC6 were “in the same boat” to deal with academic and life issues that they were facing. Given the first two term’s friendship, they could fully trust and open up with each other to share their worries or boredom. Similarly, Participant AC9 increased her online contact frequency with a Chinese friend studying at a different university in London, also because of their friendship that they knew each other before coming to the UK.

Specifically, the participants with close L2-speaking contacts (n=4) increased English face-to-face contact with their partner or cohabitants because they lived together. Because of loneliness, BI1 moved to the building where her African friends live and AC2 moved to Sheffield to stay with her British boyfriend. Participant BC8 still lived with her British ex-boyfriend in the same house and Participant AC12 lived in the same building as his closest Korean friend. Within a small bubble, they “*isolated together and did not keep distance from each other*” (AC12). They cooked, ate, and went for a walk together during the lockdown. As AC12 reflected:

We are already very close. Now we feel more dependent on each other. We feel like we are going through this together. Most of the time we speak in English. I don't have anyone to speak Chinese with, so the only language I speak is English.

It showed that for those who had strong L2 social networks before the lockdown, their relationship was strengthened due to the crisis.

On the contrary, those who stayed in the UK seemed to be in a parallel world with their classmates who flew back to their home countries (Type 4). Due to the large time difference and the different mainstream social media (WeChat for Chinese students and WhatsApp for international students), the returnees did not have much common life experience to talk with their classmates in the UK or in other home countries. As their emotional bond was weak, only very few stayed in touch occasionally with foreign friends. For example, three interviewees that returned to China at T3 only maintained contact with foreign friends on WeChat, but not on other platforms. Schoolwork was their main topic, such as doing a mock speaking exam (n=1) and asking them to

complete surveys for their research projects (n=1). Similarly, AC4 in the UK did not maintain contact with her Malaysian friend who returned to Malaysia.

For the old friends and families that had close emotional bonds but were not in the same country as the interviewees, the contact frequency was higher than pre-COVID time (Type 3). Their families usually checked in upon the interviewees because they worried about them. The interviewees might not discuss the details of daily problems with them but they did receive mental support from them. As for old friends, they exchanged news and life updates from where they were living. For example, AC4 and BC8 caught up with their old American friends during the lockdown, which maintained some sort of interpersonal interaction to combat loneliness.

As for the classmates who were physically in the UK but did not have deep emotional bonds with the interviewees (Type 2), they were bonded by similar goals such as keeping safe and completing their study. Participant AC7 viewed his relationship with other Chinese friends *“similar to before”* and *“not affected by pandemic”*: *“We have to spend time with each other more, not because we like each other more”*. They formed loose support groups on social media, to purchase flight tickets, share information, and help each other with homework. China cut off most international flights since late March, so the students had to put much effort in purchasing flight tickets.

5.2.3.3 SA experience

This part was included because COVID-19 not only exerted a profound impact on the quantity and quality of L2 contact of international students directly. By affecting the SA experience, students' mental well-being and attitudes toward SA and cultural differences subtly changed their motivation for English learning and career planning, subsequently influencing their L2 contact and social networks.

Above all, the influence of COVID-19 was long-lasting. In a globalized world, although the lockdown in the UK started in March 2020, changes in the life of participants were recorded as early as T2 in February. In January, when China started the lockdown, the

Chinese participants spent much time following news in China, staying in touch with their families and friends in China. From the initial outbreak of the pandemic, Chinese students' SA life had been very different from the rest of the international students. Many were distracted from fully enjoying their SA experience and they had been alert to the possible spread of the coronavirus and the worsening situation in the UK. Participant AC4 even volunteered in bulk purchases of masks and shipped them back to China. Since the lockdown, depression affected the efficiency of interviewees and some changed their biological clock by staying up at night and sleeping in during the day. Fortunately, some interviewees (n=3) had a positive mindset and recognized that the lockdown offered a quiet time for self-reflection and self-learning (e.g., improving English reading and writing).

A race-specific impact of COVID-19 was that it heightened the negative attitudes or hostile behaviors toward Asians, especially the Chinese, subtly making some students less willing to invest efforts in cross-cultural friendships. Before COVID-19, outright racism was uncommon. Only two interviewees brought up racism when commenting on why they did not like making English-speaking friends. They either experienced microaggressions that "*some foreigners don't like to talk to you after class*" (AC6) or felt they were not treated as equally as the locals due to the language barrier, which made her "*lack the courage to make new friends confidently*" (BC1). These examples showed that personal SA experience had an influence on the L2 social network formation of individuals that they might refrain from such interpersonal communication.

But after the COVID-19 outbreak, the scale and frequency of hate incidents increased in the interview, such as verbal abuse and harassment (n=4). These incidents can be categorized into two stages. In the first stage from February to March, before the national outbreak in the UK, the discrimination took place based on the *race*, for fear of contracting the virus from Asian-looking people. For instance, at a supermarket, a local man shouted at Participant BC7: "*Stay away from me! I want to be safe!*" because she was Chinese. Media propaganda and political manipulation of terms like "Chinese virus" or "Wuhan virus" by influential political figures misled the public and created

xenophobic sentiments. In the second stage since the lockdown in March, the virus transmission was rampant. The racism incidents the interviewees encountered extended to their *looks* (i.e., what they wear). Chinese students took this virus seriously and wore masks to prevent virus transmission while in Western countries healthy people did not wear masks. Because of panic, except for masks, Participant AC4 wore goggles a few times and Participant AC14 wore a raincoat to supermarkets to prevent infection and some passers-by shouted/laughed at them. In this situation, AC4 chose to shout back and continued wearing this protection, while AC14 just shrugged it off and decided not to wear that outfit again.

It should be recognized that all the hate incidents were initiated by strangers to the interviewees. With friends and teachers in the foundation program, they openly discussed issues such as mask-wearing (AC13) and virus origin (BC1), which did not cause conflict. Fortunately, the Chinese interviewees recognized that individuals did not represent the whole country. Admittedly, some students who were undecided about their future plans chose to take internships or work in China after graduation, instead of looking for jobs in the UK. Thus, their efforts to make L2 speaking friends might decrease as a result.

5.2.4 Summary of the post-COVID informal language contact

The total amount of informal language contact at T3 was curtailed significantly compared with T2. Due to the lockdown, the tendency for more receptive L2 contact than interactive L2 contact was intensified. The main sources of L2 contact were online classes and completing assignments. The frequency of L1 use also increased considerably. Interviews revealed that students were disappointed about the loss of English-speaking opportunities in class and with service personnel. They browsed news websites more to obtain COVID information and they indulged themselves in multimedia entertainment to kill the time they spent alone in their rooms.

As for social networks, most students' overall social circles shrank in size, and activity multiplexity plummeted at T3, against the backdrop of the “international students diaspora” and the national lockdown for COVID-19. Interview data revealed that they spoke less with people in L1 and L2. The increased online contact usually took place with their L1-speaking friends or families, while their relationship with their L2-speaking acquaintances (e.g., classmates) became more distant. At an uncertain time, the international students were inclined to open up and share their daily life issues with their closest contact in the UK, because they knew they could count on these people and they would understand their situation. Unpleasant hate incidents during the pandemic were more common than before and it might subtly hold back the willingness of some Asian students to integrate into the L2 community.

5.3 SUMMARY OF INFORMAL LANGUAGE CONTACT

It can be concluded that the foundation program students in the UK used L1 more than L2 throughout the academic year and they had higher L2 non-interactive contact than interactive contact. At T2 the total amount of L2 informal contact was comparable to T1 but some socializing time was sacrificed by heavier academic workloads. Some students were disappointed at the lack of English speaking opportunities during SA, but many felt more confident in using English to complete miscellaneous tasks. From T2 to T3, L2 informal contact witnessed a dramatic decrease due to the national lockdown. Academic L2 use (e.g., exam, IELTS) and multimedia recreation made their language contact pattern more receptive.

The social network mainly concerns interactive L2 contact of the sample, which reveals the quality of their informal language contact. It was found that foundation students usually had more L1-speaking friends than L2-speaking friends. Their social network is typically comprised of close co-national friends, with some distant international friends and no British friends. Their international friends were mainly like-minded classmates, and their limited British contacts were their teachers or romantic partners.

From T1 to T2, there was barely any change in their weak and small L2 social networks. From T2 to T3, their interpersonal contact plummeted and their online contact frequency with their L1-speaking friends and family increased. Only a few students managed to establish intimate L2-speaking friendships during their foundation year.

CHAPTER VI FINDINGS PART THREE

WORKING OUT THE RELATIONSHIP

In this chapter, the answer to Research Question 3 was explored mainly based on T1 and T2 test data:

3. What linguistic factors (corpus frequency, semantic opacity), personal factors (overall English proficiency, L1), and L2 exposure (language contact, social networks) influence phrasal verb (PV) development in these participants during study abroad?

The chapter is divided into two parts: the quantitative analysis and the qualitative analysis. Overall, the positive relationship between informal language contact and PV knowledge became stronger from T1 to T2. However, there was no strong relationship between informal language contact and PV knowledge gain. Qualitative interview data at T3 suggested that the type of language contact might influence the PV gains students make during study abroad (SA), whereby interactive contact was reported to more likely yield PV gain.

6.1 QUANTITATIVE ANALYSIS

The quantitative analysis employed three statistical measures, being correlation analysis, multiple regression analysis, and mixed-effects modeling. The complexity of these analytical methods increased gradually, from examining the relationship between individual pairs of variables to combining PV features and learner features in the same predictive model. These analyses complement each other in this chapter, to triangulate the quantitative relationship between PV and language contact. Further, correlation analysis and regression analysis mostly examined variables at the same time point, whereas the mixed-effects modeling merged data from T1 and T2 into one dataset and reveal a holistic picture of the relationship.

6.1.1 Correlation

Correlation analysis was performed to detect any easily discernable relationship between language contact and PV knowledge, and between social networks and PV knowledge. For each time point, language contact factors from the exploratory factor analysis (EFA), individual items in the Language Contact Questionnaire (LCQ), and social network variables from the Study Abroad Social Network Survey (SASNS) were tested for their correlation to receptive and productive PV scores respectively. It should be noted that a causal relationship was not inferred from this analysis and Bonferroni correction was applied to correlation analyses based on the number of analyses (LCQ items significance threshold: $.05/18 = .003$, LCQ factors significance threshold: $.05/3 = .017$, SASNS significance threshold: $.05/10 = .005$).

6.1.1a Time 1

At T1, among the language contact factors from the EFA, the only significant relationship was between L2 receptive contact and productive PV scores (Table 6.1). When it comes to individual items from the LCQ, there were two significant relationships between receptive PV scores and receptive L2 activities and four significant relationships between productive PV scores and receptive L2 activities (Table 6.2). The strongest association was between productive PV scores and the frequency of browsing English websites ($r = .394$, $p < .001$). Overall, the relationship between language contact and PV scores was weak and marginal.

Table 6.1 Correlation between T1 PV scores and LCQ factors

Language contact factors from EFA	Productive PV (N=140)		Receptive PV (N=136)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Receptive contact	.236**	.007	.210	.019

Note: ** items were significant after Bonferroni correction.

Table 6.2 Correlation between T1 PV scores and LCQ items

LCQ items	Productive PV (N=134)		Receptive PV (N=128)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Reading books, newspapers or magazines	.203	.019	.269**	.002
Browsing websites	.394**	.000	.332**	.000
Reading timetables, menus, maps, announcements	.293**	.001	.222	.012
Playing computer and mobile games	.339**	.002	.230	.049
Listening to people's conversation	.255**	.003	.242	.006

Note: ** items were significant after Bonferroni correction.

The social network variables from the SASNS did not have any significant relationship with the PV test performance in the Spearman correlation, demonstrating the slow establishment of L2 friendship (see Appendix K for details).

6.1.1b Time 2

At T2, after one term abroad, the relationship between informal language contact and PV performance became more salient. Two out of three factors from the EFA revealed significant relationships with PV test scores (Table 6.3). Reading contact and multimedia contact had moderate positive relationships with PV scores. In terms of individual items, seven items showed a significant relationship with productive and receptive PV scores (Table 6.4). Especially, multimedia activities with authentic spoken English input had moderate relationships with PV scores. Likewise, the frequency of receiving informal English written input (e.g., browsing websites; using social media; reading timetables, menus, announcements, and signs) was also positively related to PV scores. Activities entailing writing output had a weak relationship with receptive PV scores. Lastly, speaking English with international friends and listening to English songs had a weak significant relationship with receptive PV scores, but not productive PV scores. The result of correlation analysis can only establish that participants who engaged in these activities tended to have good PV knowledge, but not enough to confirm that they had a good command of PVs because of engaging in L2 activities.

Table 6.3 Correlation between T2 PV scores and LCQ factors

Language contact factors from EFA	Productive PV (N=140)		Receptive PV (N=136)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Multimedia contact	.357**	.000	.372**	.000
Productive contact	.171	.044	.193	.024
Reading contact	.401**	.000	.441**	.000

Note: ** items were significant after Bonferroni correction.

Table 6.4 Correlation between T2 PV scores and LCQ items

LCQ items	Productive (N=148)		Receptive (N=143)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Reading books, newspapers or magazines	.190	.021	.194	.020
Browsing websites	.451**	.000	.457**	.000
Reading timetables, menus, maps, announcements, etc.	.332**	.000	.420**	.000
Using social media	.395**	.000	.354**	.000
Using social media in L1	-.125	.131	-.165	.048
Watching TV	.437**	.000	.455**	.000
Watching movies	.301**	.000	.332**	.000
Watching YouTube	.369**	.000	.331**	.000
Playing computer and mobile games	.081	.328	.157	.062
Listening to songs	.179	.030	.245**	.003
Listening to people's conversation	.081	.325	.108	.200
Listening to radio or podcast	.197	.017	.215	.010
Writing emails	.241**	.003	.265**	.001
Writing homework	.212	.010	.210	.012
Speaking with L1 friends	.074	.371	.118	.162
Speaking with international friends	.160	.052	.271**	.001
Speaking with British friends	.072	.382	.099	.239
Speaking with service personnel	.087	.294	.161	.055

Note: **items were significant after Bonferroni correction.

As for the relationship between social network variables and PV scores, none of the correlations reached significance in the Spearman correlation (Appendix K). After Bonferroni correction, it was only found that the more friends the respondents provided detailed information about, the higher scores they obtained in the productive PV test (p

=.250, $p = .004$).

6.1.1c Time 1-Time 2 gain

One reason for conducting longitudinal studies on PV knowledge and language contact is to find out whether T2 language contact had a relationship with the receptive and productive PV score gains between T1 and T2. Among the language contact factors from the EFA, only the reading contact factor had a weak positive correlation with productive PV gains ($r = .282$, $p = .011$, $N = 80$). As for LCQ items, the frequency of listening to radio or podcasts at T2 had a weak positive relationship with productive PV gain ($r = .324$, $p = .003$, $N = 84$). The other non-significant relationships after Bonferroni correction are shown in Appendix K. In terms of social networks, there was no significant relationship between T2 SASNS variables and PV score gains in Spearman correlation (Appendix K). Overall, it seems that non-interactive L2 use has a weak relationship with productive PV gains based on correlation analysis. The gain in the receptive PV test was too marginal to detect any significant relationship with L2 contact.

6.1.1d Learner corpus and informal language contact

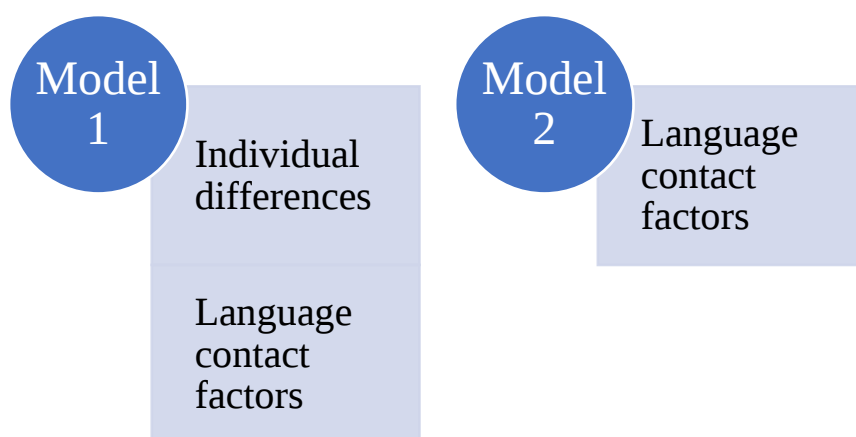
Spearman correlation was performed on language contact factors from the EFA, social network variables, and four PV variables of the learner corpus (i.e., frequency, semantic accuracy, grammatical accuracy, and sophistication) at T1 and T2. After Bonferroni correction, no variables of the PV use in the learner corpus showed a significant relationship with the language contact variables or social network variables (Appendix K). Thus, a relationship between the quantity and quality of L2 language contact and the productive ability of the participants to use PVs in speaking communication could not be established. In other words, students' spoken PV output was insensitive to the level of L2 exposure that they received. A caveat for this finding is that the sample size might be too small for enough statistical power.

6.1.2 Multiple regression

In the current study, hierarchical regression was employed to reveal the significant

language contact predictors of PV scores, after considering the personal attributes of foundation students at T1 and T2 (the sample size at T3 and that of the learner corpus did not allow for this analysis). Before the analysis, a few assumptions were checked. Firstly, the number of cases met the ideal number of $N > 50 + 8m$, ensuring the statistical power (Tabachnick & Fidell, 2013). There was an absence of multicollinearity based on correlation coefficient (independent variables correlation < 0.7) and collinearity statistics (Tolerance > 0.1 , VIF < 10). The residuals were independent, judged by the Durbin-Watson value (close to 2). The residual plot had good normality and homoscedasticity, and the values of the residuals were normally distributed (Appendix K). Lastly, influential cases that bias the model were absent, based on the value of Cook's Distance.

Figure 6.1 Two models for multiple regression



The same two models were applied to predict productive and receptive PV scores at T1 and T2 respectively (Figure 6.1). Model 1 was a hierarchical regression model which consists of a block of individual differences factors and a block of language contact factors from the EFA on top of that. Different combinations of the personal attributes were trialed (e.g., IELTS score or VLT 5k score, with motivation or without motivation). The VLT tests were administered during T1 and T2 data collection, while IELTS scores were simply reported by participants at T1. Thus, cross-validating the fit of these two variables was necessary to identify a better proxy of overall English proficiency. L1 was included as a binary variable that only distinguished Chinese participants and non-

Chinese participants. Motivation used the mean score of the Ideal L2-Self items. Model 2 was a standard regression model which only contains language contact factors in one block. The social network variables were not added to the final model because they were not significant predictors when accompanied by language contact or overall English proficiency and they would reduce the adjusted R^2 because of the large number of independent variables.

The results revealed a better model fit at T2 than at T1, which are presented below. Overall English proficiency was always a key predictor of PV scores, and VLT scores improved the model fit compared with IELTS scores. At T2 language contact factors were identified as predictors of PV scores. Thus, these findings indicate that English proficiency was the determining factor of PV knowledge, and language contact only played a role in this sample when the SA duration was long enough.

6.1.2.1 Time 1

At T1, for the receptive PV scores, the abovementioned two models were trialed. In Model 1, L1, motivation, and VLT scores were put in Block 1, with the LCQ factors (writing contact, speaking contact, receptive contact) in Block 2. The model explained 29.2% ($R^2=.292$) of the total variance of the receptive PV scores [$F(6, 114) = 7.836, p < .001$], with students' VLT score being the only significant predictor ($p < .001$) (see Appendix K). In Model 2, only the three LCQ factors were put in the same block and the model result was not significant ($p = .142$). In other words, language contact factors at T1 alone could not predict receptive PV scores yet.

In terms of productive PV scores, the result from the same Model 1 was significant [$F(6, 117) = 5.975, p < .001$] and the variance it explained was also small ($R^2=.235$). Similar to the receptive PV scores, the only significant predictor was VLT scores of the participants ($p < .001$). None of the language contact variables showed significant influence on the productive PV scores (see Appendix K). As for Model 2, the model was not valid ($p = .058$), which means language contact variables could not significantly predict productive PV scores by themselves either at T1.

6.1.2.2 Time 2

At T2, similar to correlation analysis, the variance that could be explained by the regression model increased considerably. For the receptive PV scores, in Model 1, personal factors (i.e., L1, VLT scores) were put in Block 1, whereas the EFA language contact factors (i.e., productive contact, reading contact, multimedia contact) were added in Block 2. Motivation (collected at T1) was not included in the final T2 models, because its inclusion did not enhance the model fit, suggesting that motivation was a dynamic construct that fluctuated over two terms. Model 1 explained 44.5% of the total variance [$F(5, 117) = 18.758, p < .001, R^2 = .445$], with VLT scores ($p < .001$), leisure multimedia contact ($p = .050$) and reading contact ($p = .021$) being three significant predictors. L1 and productive contact were not identified as significant predictors (Table 6.5). In Model 2 with only language contact variables (Appendix K), the result was still significant [$F(3, 121) = 12.462, p < .001$]. Though the variance that it could explain dropped to 23.6% ($R^2 = .236$). The language contact factors were all significant predictors ($p < .05$). This result suggested that at T2 the frequency of language contact itself could predict some variances of PV command, but the predictive power was stronger when taking overall English proficiency into account.

Table 6.5 Hierarchical regression results of receptive PV scores at T2

Model	B	Std. Error	Beta	t	Sig.
(Constant)	7.717	.921		8.377	.000
Chinese L1	-.810	.690	-.089	-1.173	.243
VLT2	.298	.048	.465	6.238	.000
Multimedia contact	.887	.447	.223	1.984	.050
Productive contact	-.749	.412	-.188	-1.817	.072
Reading contact	.946	.404	.226	2.341	.021

As for the productive PV score, in Model 1, the same personal variables and language contact variables were put into the model in two blocks. This model explained 43.6%

of the total variance of productive PV scores [$F(5, 117) = 18.113, p < .001, R^2 = .436$] (Table 6.6). VLT scores ($p < .001$), multimedia contact ($p = .025$), and reading contact ($p = .034$) were significant predictors. L1 and productive contact were not identified as significant predictors. In Model 2, only language contact variables were included in one block and this model could still predict 22.5% of the total variance [$F(3, 124) = 12.031, p < .001, R^2 = .225$]. All the language contact factors were significant predictors ($p < .05$) (Appendix K).

Table 6.6 Hierarchical regression results of productive PV scores at T2

Model	B	Std. Error	Beta	t	Sig.
(Constant)	4.704	.764		6.160	.000
Chinese L1	-.032	.572	-.004	-.057	.955
VLT2	.258	.040	.489	6.510	.000
Multimedia contact	.844	.370	.258	2.277	.025
Productive contact	-.652	.342	-.199	-1.908	.059
Reading contact	.718	.335	.208	2.144	.034

6.1.2.3 Other regressions

The present study is also interested in whether the PV score gains between T1 and T2 could be predicted by the T2 language contact factors, so they were put in Block 2. The same two independent variables (L1 & IELTS scores) were put in Block 1. The dependent variables were receptive PV gain and productive PV gain respectively. Neither model saw significant results ($p > .05$), which was not surprising due to the minuscule gains. Therefore, the T2 language contact could not predict the PV gains from T1 to T2.

The SASNS variables alone can predict a small variance of productive PV scores at T1 ($R^2 = .280$) and T2 ($R^2 = .158$) in standard regression models, but not for receptive PV scores. As an extra exploration, standard regressions showed that social network variables could predict some variances of the LCQ scores at T1 ($R^2 = .412$) and T2 ($R^2 = .265$), suggesting its relationship with L2 competence was indirect and might be

mediated by language contact.

6.1.3 Logistic mixed-effects modeling

6.1.3.1 Construction of the models

To understand the relationship between PV linguistic features, participants' individual differences, and their PV performance, logistic mixed-effects modeling was performed on the sample using the *glmer* function (Bates et al., 2021) from the *lme4* package in R (R Development Core Team, 2021). Mixed-effects modeling is particularly useful when the independent variables are nested in different levels and when a study adopts a repeated measures design with multiple data collection points (Winter, 2019). In the current study, the influence of PV characteristics, individual differences, and time points were analyzed in the same model. The productive test and the receptive test scores were examined respectively with the same combination of variables. The analysis was performed on the T1 dataset, T2 dataset, and the whole datasets (T1 & T2) respectively, to cross-validate the results. T3 data was not included in the analysis to enhance result reliability, given the different external environment and data collection method.

A few assumptions for mixed-effects logistic regressions modeling were checked. Firstly, the residuals had linearity and they are normally distributed. This was checked through the histograms and Q-Q plots of the models (see Appendix K). Further, the residuals should follow the homoscedasticity assumption, checked by the scatterplot of residuals (Appendix K). Secondly, there was no multicollinearity, which was checked through the correlation analysis between the independent variables in previous chapters ($r < 0.5$). Lastly, the independence of observations assumption was fulfilled because the answer to one PV question was not interfered with by the answer to another PV question.

Based on several sample size simulation studies for logistics mixed-effects models, having more groups matters more than having more cases in one group, to get high power and accuracy (Ali et al., 2019; Moineddin et al., 2007). Therefore, in the current

study, the adequacy of sample size is primarily influenced by the subject number rather than the PV item numbers. Maas and Hox (2005) pointed out that having 50 groups is inadequate and a group size of 30 is commonly acceptable in education studies. Ali et al. (2019) estimated that having over 100 groups and a group size of 50 is satisfactory. The current analysis had 160 subjects in total and 20 PV cases by each subject. Since the subject number was large, the results could be deemed fairly sufficiently powered.

Construction of the final models involves various rounds of trial and error. IELTS scores and VLT 5k scores were trialed as proxies of overall English proficiency. Similar to the hierarchical regression, VLT score was retained in the model instead of the IELTS score, for its higher statistical significance in the model, and its reliability as a direct (i.e., not reported) measure. Two variables from the SASNS were put in the model, the size (the number of English-speaking friends) and the multiplexity level (the total score of activity multiplexity and topic multiplexity) of the English-speaking social network. Since some participants did not have any English-speaking friends, the multiplexity measure was chosen to represent the quality of L2 social networks of the sample. The total LCQ score was chosen because some of the LCQ factors from the EFA were highly related ($r > 0.5$) which might cause collinearity. All the independent variables were centered based on their grand means, for easier interpretation of the intercepts in the final models (Winter, 2013).

Participant and PV items were included as two random effects and a full random slope structure was used, as recommended by Barr et al. (2013). Since the initial model was too complicated to converge, three by-subject and by-item slopes for key within-subject (LCQ scores, VLT scores) and within-PV effects (PV BNC corpus frequency) were kept in the final model. The fixed effects were VLT scores, LCQ score, L1, time point, network multiplexity, PV frequency, and PV transparency. An interaction effect between language contact scores and time point was also examined, to reveal whether the change of language contact over time played a role in affecting the PV scores. The interaction effect between PV transparency and PV frequency was examined but did not show any significance, so this interaction was removed from the final model. The

final model is: PV score $\sim$ LCQ score + PV frequency + VLT score + time point + L1 + multiplexity + PV transparency + LCQ score * time point + (VLT score + LCQ score | PV item) + (PV frequency | Subject).

6.1.3.2 Results of the models

The dataset for the receptive PV test included 3973 cases from T1 and T2. The most significant predictor was VLT scores ($z=6.885, p<.001$), followed by PV frequency ($z=3.262, p<.001$), LCQ scores ($z=3.198, p=.001$), and the interaction of LCQ scores and time point ($z=2.044, p=.041$) (Table 6.7). To be specific, there was a significant positive effect of better overall English proficiency, higher PV corpus frequency, and higher language contact frequency on the receptive PV scores of the participants. As shown by the interaction effect, the influence of language contact was stronger at T2 than at T1.

Table 6.7 Model summary of the mixed-effects model for receptive PV scores

Fixed effects	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.380	0.316	1.202	0.229
VLT score	0.090	0.013	6.884	5.83e-12 ***
PV frequency	0.026	0.008	3.262	0.001 **
LCQ score	0.023	0.007	3.198	0.001 **
LCQ score*Timepoint	0.020	0.010	2.044	0.041 *
Timepoint	0.045	0.103	0.437	0.662
L1	0.062	0.188	0.328	0.743
Network multiplexity	-0.004	0.003	-1.270	0.204
PV transparency	-.0793	0.543	-1.461	0.144

As for the productive scores, the dataset included 3680 cases from T1 and T2. The significant predictors were: VLT scores ($z=5.095, p<.001$), PV frequency ($z=3.670, p<.001$), time point ($z=2.765, p=.006$), and LCQ scores ($z=2.617, p=.009$) (Table 6.8). In other words, higher overall English proficiency, higher PV corpus frequency, longer stay in the SA environment and higher frequency of L2 language contact had significant effects on the productive PV scores of the sample.

Table 6.8 Model summary of the mixed-effects model for productive PV scores

Fixed effects	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.751	0.424	-1.774	0.076
VLT score	0.102	0.020	5.095	3.49e-07 ***
PV frequency	0.044	0.012	3.670	0.000 ***
Timepoint	0.358	0.129	2.765	0.006 **
LCQ score	0.024	0.009	2.617	0.009 **
LCQ score*Timepoint	0.017	0.012	1.409	0.159
L1	0.137	0.234	0.583	0.560
Network multiplexity	-0.001	0.004	-0.261	0.794
PV transparency	-1.184	0.833	-1.421	0.155

6.1.4 Summary of quantitative findings

Overall, a few key findings can be summarized from quantitative analysis. Firstly, correlation analysis showed that there were significant relationships between the frequency of non-interactive L2 contact and PV scores, while the productive language contact seemed too low to demonstrate such a relationship. Secondly, hierarchical regression revealed that the strength of relationships between language contact and PV scores was stronger at T2 than at T1, as language contact alone was not sufficient in predicting PV scores at T1. Thirdly, when all factors were taken into consideration together in mixed-effects modeling, high overall English proficiency and high corpus frequency of the target PVs were the most important predictors of students' PV performance, while L2 engagement and lengths of stay played a minor role. Lastly, the social network variables of participants were generally too weak to show a significant relationship with PV performance.

6.2 QUALITATIVE ANALYSIS

In this section, qualitative data will be analyzed to reveal whether and how informal language contact is related to the development of PV knowledge during SA. At T1 and T2, to prevent *the priming effect*, interview questions were targeted at general

vocabulary and phrases, without mentioning PVs in particular. At T3, students were asked questions with a focus on PVs (see Appendix G). Therefore, T3 interview data will be mainly drawn on in this analysis. The section is structured from two aspects, namely, the students' attitude toward PVs and their perceived relationship between informal language contact and PV development during SA.

6.2.1 Attitudes toward phrasal verbs

6.2.1a Attention to PVs

Overall, Chinese students rarely paid attention to PVs (n=7). For the high-frequency PVs, students reported knowing them already and having heard them all the time without a need to pay attention, suggesting automatization of these PVs. As Participant AC17 described: *"Sometimes we use it without realizing we use it"*. For the PVs that they did not know, if they were made up of simple words, they could roughly guess the meaning from the context. Their attention was usually drawn to complicated one-word vocabulary (especially nouns) because they *"have no way to guess its meaning"* (AC16). In contrast, two non-Chinese interviewees (BI1, BE1) said they did notice PVs in daily life. Participant BE1 commented: *"I feel that it's like a puzzle. You are trying to figure out what is this, if I use 'come on' and 'come in', what's the difference? I like puzzles!"* It showed that different learning styles might influence what type of vocabulary was learned by the participants. Chinese participants seemed to focus on one-word vocabulary while other international students might pay more attention to colloquial vocabulary such as PVs, which aligns with the finding on VLT gains in Table 4.15.

As for the modality from which learners tended to notice PVs, written form seemed to be easier than aural form (n=4). Reading was a typical way for students to learn new vocabulary, including PVs. As Participant AC9 put it: *"I notice PVs in books, when I finish some exercises. I have the habit of underlining the phrases in books, or papers I read. I will try to remember some expressions that look special. If I just hear it in daily life, I might not notice"*. It showed that PVs with 'special' meanings or forms were more

salient to be noticed. Participant AC14 explained why she missed out on PVs in people's conversations: *"If I hear something, I pay more attention to the overall meaning of a sentence; If I understand the main idea, I won't care too much about the details in this sentence"*. Interpersonal conversations had high instantaneity, during which a pause to ponder on a PV was impractical. It seems that some Chinese interviewees had relatively weak listening competence, which made it hard for them to pick up PVs in speaking and make use of the L2 exposure during SA. This finding corroborated the significant relationship between reading contact and PV scores and the non-significant relationship between speaking contact and PV scores in Section 6.1.1.

6.2.1b Usefulness of PVs

Students unanimously agreed that PVs were useful for daily communication (n=20). Firstly, they recognized that local people used PVs a lot and PVs made spoken discourse more native-like and colloquial. Participant AC9 believed that *"to some extent, they [PVs] are more useful than memorizing the sophisticated words"*. Because the students *"not only study here but also live here"* (AC7), they had a pragmatic need of understanding English native speakers. Secondly, some students found PVs useful, because they could express meanings such as *"people's emotions and actions"* (BC8) in a concise way, which *"is easy for others to understand"* (AC3). These excerpts showed that students recognized the function of PVs in facilitating spoken communication. Six participants pointed out that PVs were more useful in oral English, but not in academic writing. For those whose main goal was just finishing their assignments in English, it was admitted that *"we don't need to learn a lot of PVs. It's too advanced for me"* (BC4). In contrast, Participant AC9 had advanced English proficiency and she thought that *"even in IELTS exam, examiners would like to hear you use them [PVs]"*. This excerpt showcased that the interviewee was aware that the ability to use colloquial expressions was desirable for advanced L2 learners. Overall, this finding revealed that the academic prioritization and lack of L2 speaking opportunities lead to the low importance attached to PV acquisition.

6.2.1c Difficulty of PVs

Interviewees gave mixed opinions on the perceived difficulty of PVs. On one hand, they (n=9) recognized that PVs were easy. Three interviewees reflected that they did not do well in the test because they forgot many PVs that they learned at school. Students were familiar with high-frequency PVs (e.g., go out, come on), as noted by Participant BC8: *“Because I can use it daily; I use it regularly; I will speak it in my daily life”*. For the PVs that were not taught to them before, if they were *“made up of simple words combination; you will guess the meaning when you see them”* (AC13). On the other hand, some students (n=7) considered PVs difficult and five students regarded PVs as “confusing”. The difficulty lies in a few aspects, such as their infrequent presence in their L2 exposure. The figurative meaning and polysemous nature of PVs also increased the difficulty in using them: *“I think two words appear together having completely different meanings. I think it’s difficult to remember so many meanings and I easily forget them. I have to rely on experience. I can’t remember them in rote-learning”* (AC13). Further, they tended to pay more attention to one-word vocabulary than phrases in their learning, so their PV knowledge tended to lag behind the one-word vocabulary.

A few interviewees (n=8) pointed out the importance of practice in mastering PVs, which lends support to usage-based theory. For example, Participant BE1 linked PV difficulty to the frequency of use, in observing that:

I think they are not difficult or easy. It’s not about difficulty. It’s about practice: how often do you practice PVs. If you are practicing in your daily life, it becomes normal. It’s difficult for people who don’t use it or know how to use it, so it became more confused. “What’s the meaning of this?” They don’t realize this word and this word [together], the meaning will change! You cannot even realize the difference.

It can be seen that the challenge of learning PVs mostly comes from the fact that many participants did not use PV often. The challenges of polysemy, semantic opacity, and

confusing particle choice could be tackled by sufficient colloquial L2 exposure. If the learners use PVs often enough, the suitable PV choice will be automatized. In self-learning, it is hard to recall the meaning of PVs without a context of communication. This finding supported the result from mixed-effects modeling that PV frequency overrode PV semantic transparency in predicting PV knowledge.

6.2.1d How to learn PVs

When being asked whether they were more likely to learn PVs from class or daily life, most students chose daily life (n=12). Participant BC6 explained: *“You listen, you learn it and you use it... We just use it [PVs] in our daily life more than in class, so I think it’s very important to learn in daily life”*. When they heard others using PVs, they could learn the context and meanings of the PV, which is crucial for their productive use of the target PV. The context played an indispensable role in facilitating PV acquisition, as illustrated by Participant AC13: *“I need to hear other people using it, then I will realize: ‘Oh I can use it this way.’”* However, since most participants did not speak English often in daily life, their PV gain was limited despite living abroad.

Nevertheless, some interviewees with lower English proficiency preferred to learn PVs from class (n=4). They needed explicit instruction and formal teaching to acquire new PV, as explained by Participant BC7: *“I think I learn PVs from class; I need to memorize it well before I can use it fluently in life; If I haven’t learned this vocabulary formally, I can’t produce it; I have to count on teachers”*. These learners were used to the traditional way of learning with textbooks and PowerPoints and taking notes to memorize target phrases. This disparity showed that less proficient participants either rarely had meaningful spoken interaction in English after class, or could not understand/remember new exemplars of PVs even if they heard them in daily life.

Lastly, some interviewees believed that both sources of learning were useful (n=4). For example, Participant AC14 believed that *“in class, you will be exposed to many English expressions. In daily life you can see how this PV is used”*. In-class instruction is the

premise and the repeated exposure deepened her understanding of how to use the PV in context. Besides, Participant AC2 raised a point about the importance of attention in their uptakes: *“I think both. It really depends on your focus. No matter you are listening to class or you are talking with your friends, if you have high focus during that time and you will listen the key points from what I say, you'll definitely remember it”*. This excerpt highlighted the role of noticing in SLA (Schmidt, 1990).

Overall, the interviewees recognized that PVs were more common in daily life than in class. When hearing other English speakers using a PV, it is easy to learn it from the context and remember the nuances of its meaning. Since the sample has limited English proficiency and low L2 speaking contact, learners without L2 friends or with limited L2 listening competence needed explicit PV instructions to achieve acquisition. Additionally, learners' attention was important in the incidental acquisition process.

6.2.2 Phrasal verb learning in study-abroad

6.2.2a The relationship between informal language contact and PV development

Most interviewees (n=16, 73%) believed that SA was beneficial to PV development and eight of them explicitly attributed the benefit to authentic L2 exposure and the opportunities of speaking L2. Two participants believed that their English saw well-rounded improvement during SA, including PV knowledge. For example, Participant BE1 shared that, in her home country, their English learning is imbalanced because they learn academic words in English for study and work but rarely learn general words for daily life. In the UK she *“has chance to know more about the language, to know about more faces of the language”*. PVs constitute a ‘face’ of English whose natural use is often inaccessible in the EFL learning environment but available in the SA context. Similarly, Participant BC1 believed that, by talking to local people, she could understand the real meaning and usage of words whose meanings were perceived differently when she was in China.

The L2 interaction with friends was beneficial because of the diverse topics. For example, Participant AC2 found talking with people from different countries particularly helpful, because *“they learn from their daily life, the book or the article they read, and when they speak out, if you don’t understand, you will remember what they say, and you will learn it”*, which recognized the value of speaking English with non-native speakers. Secondly, L2 exposure from strangers was also beneficial. For example, *“the worst condition is that this person doesn’t have any foreign friends, but the person is still in the English environment; They will meet foreigners; They will eat in the restaurants, so they will still hear a lot what the local people speak”* (BC8). Given enough frequency, they might unconsciously become familiar with the high-frequency PVs. On the occasions that the participants had to communicate with strangers in English, it is possible for them to notice unfamiliar PVs and accumulate their usage in their mental lexicon. As AC4 reported:

Before, I don’t know what ‘sold out’, ‘run out’ means. When I go to shopping at grocery store, that’s the place where I learn those phrases. I can hear them often [during the lockdown]: ‘They run out of the toilet paper.’ ‘They sold out the paper towel’.

Moreover, the speaking opportunities in the SA context enabled participants to produce and automatize the PVs that they had learned passively. Learners could associate authentic English using scenarios with specific meanings of PVs, which is a sharp contrast to learning in the EFL context. As AC6 compared:

I think SA can improve your productive use [of PVs]. At least SA makes you more familiar with it. In China, you might see it and write it down. But when you speak it, use it, it’s definitely different from when you are in China and only see it and write it down. You will have deeper impression. You will be more fluent and more proficient.

This excerpt echoed the quantitative finding of larger productive gain than receptive

gain in PV tests, which suggests that learners could recall more PVs after SA. Her point resonated with the output hypothesis (Swain, 1985) and highlighted the importance of using PVs to internalize them effectively.

When being asked how long of stay is needed for the PV gain to take place during SA, interviewees recognized that it was not SA itself that automatically benefited PV acquisition. It was the opportunity of interactive contact with English speakers that contributed to the improvement of PV knowledge. The participants emphasized that spending sufficient time with other English speakers in SA and investing enough effort in the experience were crucial for gains in PV knowledge to take place:

I don't think the PV development is proportional to how long you stay abroad. It depends on the frequency you contact English speakers. If you spend a lot of time with British people, your English will develop fast, including PVs. Especially when you hear the teacher or your classmates use it, you will have deeper impression (AC3).

A wholehearted immersion in the English environment accelerated PV learning, whereas strong reliance on their L1-speaking social circles did no good to spoken English development. Unfortunately, many foundation students lacked the opportunity or English proficiency to establish enduring relationships with English speakers, which may have limited the rate of their PV gains. Against this backdrop, some participants viewed reading L2 contact as a way of learning too, as illustrated by the following view: *"If I learn without the text, my improvement will be slow; My English level is limited, I can't grasp the vocabulary that native speakers say quickly, so I need the books"* (AC9). She considered reading as an additional source of language contact, which was a solution to counterbalance her limited listening comprehension ability.

Lastly, three interviewees thought SA did not necessarily lead to PV gain, because they did not notice PVs: *"I think words related to daily life I will accumulate more; I don't pay attention to PVs in particular"* (AC14). What students were likely to accumulate

in SA were nouns that directly related to their study or daily survival communication, but not PVs. Unlike nouns, PVs are replaceable by one-word verbs, so knowing them is not a must for their daily life. Another point is that negotiation of meanings did not take place often, because people could guess what they meant, as reported by AC16: “*I think speaking English in the UK, you don’t have to be accurate as long as you can communicate with others*”. If students used a PV incorrectly, the interlocutor might carry on without pointing it out. After they build robust friendships with English speakers, they might feel more comfortable asking for clarifications or being corrected by their friends, thereby learning new meanings of PVs.

6.2.2b PV learning examples

Participants were explicitly asked about new PVs they acquired during SA. Overall, students did not recall many new PVs that they learned during SA, but they recalled some idioms and words that they learnt in daily life. Some PVs examples they claimed that they learned from movies might be acquired before they studied abroad (e.g., *get off, come on*). Two valid examples are *turn out* and *boot up*. AC2 tried to remember the idiomatic phrases used by her English-speaking friends and she learned *turn out* which is often used in “*it turns out to be ...*”. Another example *boot up* means something starts to work. AC4 learned it from an English learning channel in a mobile application. She shared that:

[If I memorize vocabulary list], maybe this moment I will remember it, but it won’t stay there long. If I learnt the word from podcast, I will remember the word forever. Because they do not just explain the word, they will give you the example and when you will use it. They will read short news for you.

These two examples again demonstrated how spoken input contact facilitates PV acquisition. AC2 took advantage of her close-knit international friends’ circle to learn native-like expressions, while AC4, disappointed at her lack of English-speaking social network, made up for it by listening to English podcasts. Authentic interpersonal

contact was favored because it required less self-discipline from learners than listening to podcasts. When spoken contact was not available, multimedia L2 contact could also be beneficial if the input has clear contexts and rich explanation.

6.2.2c Reflection on LCQ scores and PV scores changes

The interviewees were told their LCQ changes and PV scores gains between T1 and T2 at T3 interview and they were asked to comment on any changes in their performance to shed light on the issues that students saw as crucial to PV learning. The findings will be presented from two aspects: PV test-taking experience and SA experiences, which focus on the immediate micro-level and long-term macro-level reasons.

Repeated testing, as a form of exposure itself, had different effects on students. BC4 (receptive test: +1, productive test: +1) noted a higher level of uncertainty: *“If I do a test twice, I will think about what I did last time and I will change my answer, so the second time might not be good”*. It showed that some PVs were confusing for learners and their knowledge was not fully consolidated. On the contrary, Participant BC6 deemed the receptive PV test a useful hint in helping her acquisition from T2 to T3 (receptive test: +3, productive test: +2):

I think my short-term memory is good. There are many PVs I don't know. I learn new PVs by doing this test... I vaguely remember the answer I wrote in the first test, but it doesn't mean I am correct, so I still make some mistakes. This time my gain is bigger maybe because I took the test three times. The experiences I gained in the first two tests led to gains.

It seems that some partial knowledge of PVs was strengthened by taking the test seriously and students retained some memory of the PVs in their mental lexicon due to high concentration during the test. Most participants were not influenced by longitudinal testing, as AC4 shared: *“Every time I do the test, I don't try to recall what I answered last time; I choose the answer that I think sounds most natural and familiar”*. Lastly, some participants did not do well because of non-academic factors, such as

period pain, other sicknesses, or distraction by classmates (n=3).

Regarding macro-level SA experience, L2 exposure was still the keyword. Four students attributed the lack of PV improvement to the fact that they did not care about vocabulary or did not use English often. For example, BC3 (receptive test: -2, productive test: -1) paid less attention to English learning and focused on subject courses at T2. In Term 1 she had English classes every day and she made some foreign friends. In Term 2 she had fewer English classes per week and her two foreign friends had left the UK, so she used English less than T1. In stark contrast, three students with strong English-speaking social networks throughout the tracking period developed their PV unconsciously or ‘effortlessly’. BI1 made considerable gains in PVs from T1 to T2 (receptive test: +4, productive test: +4) and she reflected:

To be honest, it's not because I am studying [English] that this happens. I think it's just English has become a part of me. I speak English every time, every day. They just come naturally. I don't wanna use the word "naturally", but it's becoming easy for me nowadays. For me, it's kind of obvious: I don't study English at all. It's because of my friends and we speak a lot.

This finding showed that speaking with international classmates in the foundation program was beneficial for PV learning. In the same vein, AC12 (receptive test: +1, productive test: +4) attributed his gain to speaking with international friends: *“I luckily made a group of [international] friends right from the beginning, which helps me keep a high level of language contact; I spend time with them, which makes me almost reach my upper limit of English contact”*. These interviewees immersed themselves in the L2 speaking environment and reported that their PV gains took place ‘unintentionally’.

When students had a small English-speaking social network, the PV gains mainly appeared when accompanied by strong self-discipline in English learning. IELTS test seemed to be a crucial incentive for participants (n=3). For example, AC13 improved PV scores (receptive test: +2, productive test: +4) at T2 despite low LCQ scores. In the

meantime, she improved her IELTS from Overall Band 6.0 to 7.0 in half a year. She explained:

I think I tried really hard [to prepare for IELTS]. I practiced speaking and writing. Doing your survey makes me realize there are many ways to increase my English use, so I made some foreign friends. I try to create an English-learning environment for myself.

She practiced English speaking for the exam purpose and unexpectedly saw large gain in PV as well. It should be noted that some participants mentioned that they benefited from both interactive and non-interactive L2 contact in improving their English (n=2). Participant AC4, who only had a Malaysian friend to speak English with, was conscious about improving her English and often listened to English-learning channels. She witnessed PV gains at T2 as well (receptive test: +2, productive test: +4). Unfortunately, the participant that made the largest gain dropped out from the T3 data collection, so it was unknown how he accounted for the changes in his test performance.

6.2.3 Summary of qualitative findings

The purpose of the interview was to find out how foundation program students accounted for the relationship between PV learning and informal language contact during SA. In brief, recognizing the usefulness of PVs, the participants rarely paid attention to PVs and did not consider them of immediate importance. Their prioritization of academic outcomes influenced their emphasis on English learning. They considered PVs not suitable for assignments and only useful when talking to native speakers whom they rarely met. They believed the difficulty of PVs depended on the frequency of usage in daily life. Overall, they emphasized the importance of spoken communication during SA in acquiring PVs, although many of them did not speak English often. The interview data were in line with the quantitative findings in that familiarity with high-frequency PVs and the importance of language contact were recognized by learners.

CHAPTER VII DISCUSSION

In Chapter II, a comprehensive literature review shed light on the state-of-art of study abroad (SA) research, especially the relationship between language contact and formulaic language. Following this line of research, this study sought to unravel the relationship between informal language contact and phrasal verb (PV) knowledge of foundation year students in the UK. This discussion chapter will be presented by research questions (RQ) in the order of PV command, language contact pattern, and the relationship between these two constructs.

7.1 Phrasal verb development

7.1.1 Developmental profile

RQ1: What is the developmental profile of PV knowledge among international foundation program students in the UK over one academic year?

7.1.1.1 Low PV command

As illustrated in Chapter Four, the foundation students had limited receptive (54% of the full score) and productive knowledge (40% of the full score) of high-frequency PVs tested in this study. With a closer look at each PV item, high-scoring items in both receptive and productive tests (mean > .80) are PVs related to daily life actions, such as *come in*, *go out*, *go on*, which have a high occurrence in the classroom and everyday life of the participants. Research on one-word vocabulary shows that for very frequent words (up to the 2,000-word frequency level), receptive and productive vocabulary knowledge tend to be equally strong (Dóczy & Kormos, 2016). The threshold for such parallel development of PVs is unknown yet. But the current study provides a hint that this sample tends to master the Top 1-25 frequency range in the PHaVe List (Garnier & Schmitt, 2015) at both receptive and productive levels.

Due to the employment of different tests, students' PV performances in different studies are not directly comparable. But it could be concluded that learners' test scores are closely associated with learner level, test format, and the tested PV meanings. As shown in the literature review, among three existing studies employing standardized tests, the most advanced university student cohort scored the lowest (40.6% of the total score) (Garnier & Schmitt, 2016), while the youngest learner cohort scored the highest (75%) (Kamarudin et al., 2019). The participants in Schmitt and Redwood (2011) scored 65.2% on the receptive test and 48.2% on the productive test, higher than the current sample. Overall, it seems that learners tend to perform better in the receptive test which tests one meaning from each PV and had fewer distractors with very different meanings.

Compared with these tests, the PV tests in the present study are relatively difficult for the level of participants. However, referring to the large gains and the close-to-ceiling post-test results in some previous studies (Conroy, 2018; Schmitt, Dörnyei, et al., 2004), the present test was designed to account for the possible considerable gain during one academic year, mitigation of *the practice effect*, and prevention of *the ceiling effect*. Freed (1998) cautioned that if an instrument was unable to capture the progress made by more advanced students, researchers might falsely conclude that lower-level students made the greatest gains abroad. This study ascertained that intermediate English learners in general made slow gains in PV knowledge despite living abroad for seven months, which has an implication for the PV test design in future longitudinal studies.

7.1.1.2 Slow developmental profile

The PV gains made by the current sample are smaller than the gains made by the Chinese participants in Conroy (2018) with similar L2 proficiency in a short summer program. Two reasons might lead to the contradiction. Firstly, the participants in Conroy (2018) were summer school homestay students who could enjoy colloquial L2 interaction without much academic workload, whereas less than 2% of the current sample lived with local British people, limiting their opportunity of receiving L2 spoken

input. Secondly, among different aspects of word knowledge, Schmitt (1998) found that spellings are acquired quickly and mastery of their derivational forms or meanings is gained later on, and often only partially. Conroy (2018) tested students' ability to hear and transcribe idiomatic PVs which were commonly used in daily life, while the current study tested form-meaning mapping of general PVs, which might require more conscious noticing of PV items and more explicit scaffolding in the daily exposure. More tailored item selections for the sample (testing items that often appear in coursebooks/on campus) might capture the fine-grained changes in PV development.

The growth of formulaic language tends to be slow and uneven, as Laufer & Waldman (2011) pointed out for collocation development. Crossley et al. (2013, p. 743) argues that verbs are “more abstract, and ambiguous, and thus, less concrete, imageable, and meaningful”, and hence their acquisition is more strongly aided by contextual diversity. It may take a longer time and a higher level of immersion for students to acquire and memorize new exemplars of PV incidentally. In a study of vocabulary development, Dóczy and Kormos (2016) point out that vocabulary growth is slower if learners are expected to acquire new words from input alone, rather than being taught explicitly. Their evidence came from international EAP students in a British university which resembled the current sample. They found that the incidental learning conditions in this EAP program and target language environment led to variable outcomes. It seemed to benefit a smaller group of students to a larger extent, while a great majority of the students gained from it minimally (Dóczy & Kormos, 2016). Overall, it can be speculated that the gain of PVs tends to be incremental in either the SA setting or the EFL setting (Garnier & Schmitt, 2016).

Based on interview data, this study hypothesizes a hierarchy of importance of vocabulary acquisition during SA. It should be noted that this order only applies to degree-seeking international students abroad and the order of importance does not represent the order of acquisition difficulty. Rather, the order of importance reflects the frequency of usage/encounters. The more often learners need the target vocabulary in the SA environment, the more likely they will acquire the target item. The examples

below were recalled by participants in the interviews. Above all, technical/academic terms in the coursework are accumulated with the highest importance, because the participants have to use this vocabulary to understand the lectures and complete assignments (e.g., “attachment” in psychology, “stakeholder” in principles of management). The second most important type of vocabulary are nouns of objects related to daily life, especially names of daily consumer goods (e.g., detergent) and food (e.g., avocado). Students need to know these words to purchase what they need or what they like to eat.

The importance of acquiring idioms and other types of formulaic language is the lowest, which is like ‘icing on the cake’. As Wulff (2018) points out, adult L2 learners tend to employ prefabricated language less than L1 users and are less reliant on them. In fact, as formulaic language may play a lesser role in L2 acquisition compared to L1 acquisition, it is likely to matter more to some learners than others and is, therefore, best seen as optional to achieve nativelike idiomaticity rather than a strictly required route for L2 acquisition (Wulff, 2018). The participants usually accumulate formulaic language incidentally. Idioms (e.g., out of my league, spill the beans) or pragmatic routines (e.g., Anytime.) with high prominence or specific pragmatic functions are more likely to be acquired. Due to the low importance of PVs in vocabulary learning, the participants did not witness large gains without frequent interpersonal L2 communication.

7.1.1.3 Productive gain and receptive gain

Throughout the tracking period, learners’ receptive PV size was always larger than their productive PV size, in line with general vocabulary literature (Laufer & Paribakht, 1998; Webb, 2008). It was rather surprising that they made larger productive gains and just marginally significant receptive gains from T1 to T2. This finding contradicts a long-held belief that it was easier to observe receptive vocabulary gain than productive vocabulary gain (Pellicer-Sánchez, 2019). A possible reason is that informal interactions in SA are largely communicatively driven and meaning-focused, which will

more likely drive proceduralization and potential automatization of declarative knowledge, while instructed contexts tend to favor the acquisition of declarative knowledge (McManus et al., 2021). Compared with other SA literature, the current participants made 5% relative gain in the receptive PV test and 17% relative gain in the productive PV test. It seems to offer evidence that lower initial competence in certain linguistic area has larger room for the improvement to be achieved and measured during SA. The receptive PV competence of the participants was probably sufficient for their daily use, so it was harder to make gains from a higher threshold. This result echoes a common finding in other SA studies that lower-level students make the most advances because they have the most to gain (Regan, 1998).

Schmitt, Dörnyei, et al. (2004) lend support to the current study. 94 Pre-sessional English course students made larger productive gain (3.20) than receptive gain (2.03) in three months in a formulaic sequence test. At the item level, gains from receptive knowledge to productive knowledge were witnessed in 233 cases, while gains from unknown to receptive state or productive state were observed in over 50 cases. In the present study, gains from partial knowledge to full knowledge were also more common than gains from the unknown state to partial or full knowledge (Table 4.13b). It demonstrates that the benefit of SA is not limited to the acquisition of new vocabulary, but also the consolidation of partial knowledge of vocabulary (Schmitt, Dörnyei, et al., 2004). Except for large-scale research, a case study also supports this argument. Li and Schmitt (2009) investigated the lexical phrase development of one Chinese master student's assignments at a British university over one academic year. It was also found that she demonstrated increased mastery of known words rather than acquiring new words in her writing. To be specific, her confidence in using the right phrases and the appropriacy of usage were enhanced, while the frequency and the diversity of phrases did not increase.

The small gains in the receptive PV test should be interpreted with two caveats in mind. Firstly, the test format exerts a direct impact on the performance of learners (Schmitt, Dörnyei, et al., 2004). In other words, whether learners get higher scores in receptive

or productive tests to some extent depends on the relative difficulty of the instrument (Waring, 1998). Although methodological steps were taken to ensure the test measure truthfully represented the level of learners, it should be recognized that the current receptive test with eight distractors is more difficult than traditional multiple choices questions and particle selection is particularly challenging for the participants. Another factor is that learners' test performance might be influenced by motivation. Literature found that students' different scores in group administration and individual administration might be due to a lack of test-taking efforts rather than a lack of knowledge (Coxhead et al., 2018). The receptive test was administered at the end of the whole quantitative data collection when learners might feel distracted and tired. Although they were given enough time to complete the test, many may have experienced test fatigue and did not spend time checking their answers toward the end. Therefore, this psychological factor may have had an impact on the results yielded.

7.1.1.4 Learner corpus

In the attempt of analyzing PV production in a learner spoken corpus, quantitative analysis revealed that their PV usage did not see measurable gains in terms of frequency, semantic accuracy, grammatical accuracy, and sophistication. In other words, seven months abroad (including the lockdown) did not change their PV usage in spoken discourse, which implied a somewhat limited PV input in their everyday interactions. Existing longitudinal studies on formulaic language corpus generally showed that the gains pattern was non-linear and subtle. On one hand, the quantity of formulaic sequences does not necessarily increase. For example, the lexical bundle use in 15 Chinese learners' essays over one academic year in China was limited and showed a U-shape trend in frequency (Zheng, 2016), while the frequency of n. + adj. collocation in the essays of 36 Chinese learners of Italian did not change significantly after five months in Italy (Siyanova-Chanturia, 2015). On the other hand, positive changes in the quality of formulaic language production were observed in studies with long tracking periods. Siyanova-Chanturia (2015) noted more native-like and less unnatural collocations from her sample. Crossley and Salsbury (2011) found that six learners

improved their English bigram (e.g., *you know, I think*) speaking accuracy over one year. Qi and Ding (2011) found that 56 Chinese learners made gains in terms of frequency, accuracy, and diversity of formulaic sequences in oral monologues over a three-year period. It is likely that the use of formulaic language takes an extended period of time to progress, and the examined period in this study may not be long enough to capture such an incremental growth (Zheng, 2016). There was just a three-month gap between T1 and T2 data collection and the current sample entered the lockdown after T2, so, understandably, foundation students did not make significant gains in their spoken PV usage in the IELTS speaking test.

7.1.2 Influential factors

RQ1a: What are the non-exposure factors that influence the PV scores of the participants?

A few non-language-contact factors were explored in terms of their relationship with PV test scores. It is unsurprising that overall English proficiency and corpus frequency have a significant relationship with PV scores, as this echoes previous work (e.g. Qian & Lin, 2019; Schmitt & Redwood, 2011). In the literature review, there are incongruent evidences on whether low-proficiency students or high-proficiency students make larger L2 gains during SA. The current study showed that more proficient learners made slightly larger gains in the productive PV test between T1 and T2 (Table 4.14), suggesting that formulaic language knowledge is a relatively advanced L2 competence that is not easy to improve for everyone.

Semantic opacity and L1 did not show significant relationships with PV scores, which is worth discussion. Firstly, as shown in Section 2.2.5a, existing evidence on the role of semantic opacity has been mixed (Garnier & Schmitt, 2016; X. Zhang & Wen, 2019). It seems that the core issue is how semantic opacity is defined. The subjective ratings in X. Zhang and Wen (2019) seemed to be influenced by personal interpretations of the difficulty of PV meanings. The present study corroborates the findings from Garnier

and Schmitt (2016), possibly because the same researcher-determined judgment was adopted. This study supports the argument that corpus frequency overrides semantic transparency in influencing PV scores (Garnier & Schmitt, 2016). PVs with higher frequency in daily life or in formal instruction were more likely to be acquired by learners. If learners have stored the target PV in their lexicon, they could recognize or produce it regardless of the semantic opacity. On the contrary, if learners do not have an impression of the target literal PV, they would fail to come up with the correct PV by themselves even with the hint in the tests.

The second issue to be discussed is the role of L1 in PV acquisition. Existing studies have revealed that there is a strong cross-linguistic influence on PV learning (Kuiper et al., 2009). PV avoidance was observed in learners with different L1s such as Arabic, Chinese, and Persian (Becker, 2014; El-Dakhs, 2016; Houshyar & Talebinezhad, 2013), while Dutch, Polish and German learners produced similar or more PVs compared with English native speakers (Waibel, 2007). The current study did not reveal a significant L1 difference in the PV knowledge between participants from China and the rest of the cohort, due to two possible reasons. Firstly, participants in the non-Chinese group mostly came from other Asian countries (e.g., Korea, N =11; Saudi Arabia, N =5) with a handful from African countries whose L1s were also quite distant from English. Hence significant difference was not detected among themselves. Another possible reason is that merging all non-Chinese learners into a category might obscure some nationality differences in the PV test. Due to the nature of the sample, it was not possible to recruit a similar number of students from each L1. To sum up, existing evidence indicated that students in the sample had a similar level of PV command regardless of their nationality. Whether they had a relative disadvantage in PV compared with European students needs further investigation.

7.2 Informal language contact

7.2.1 Language contact patterns and developmental profile

RQ2: What are the patterns and developmental profile of informal L2 contact among international foundation program students in the UK over the first six months of studying abroad?

7.2.1.1 Language contact

7.2.1.1a Language contact pattern

Many studies on SA participants have found more frequent L1 use than L2 use (Kinging, 2008; Trentman, 2013), while some have reported the opposite findings (García-Amaya, 2017). Overall, the present study found that most participants used L1 much more than L2 in their spare time, due to their limited English level and strong co-national social circles. This finding aligns with existing research and contradicts a common belief held by the public that SA can be equated with an immersive L2 learning environment. Students within the L2 community can be more or less willing or able to exploit the opportunities that exist (Foster et al., 2014) and this phenomenon is especially common among Asian students studying in Anglophone countries (Wright & Schartner, 2013).

The present study focuses on L2 but not L1 contact. What stands out from the data is the high proportion of academic L2 contact. Quantitative data revealed that the frequency of writing homework in English was always ranked the top among all L2 using activities after class. This finding is unique compared with existing literature which found more interactive L2 contact than receptive L2 contact in SA learners (Köylü, 2016; N. Taguchi, 2008). For example, although writing homework in Spanish was ranked the second most frequent activity by American exchange students in Spain, students spent twice as much time speaking Spanish with local people (Hernández, 2010). In other studies, the contrast was more dramatic. American students studying in

Egypt spent the highest amount of time listening to Arabic and the least time writing homework in Arabic (Trentman, 2013). As Gautier (2019, p. 231) interpreted, American exchange students were “fairly relaxed” on the stakes related to SA and considered it “a year of vacation”, while Chinese learners pursuing degrees abroad considered SA a high-stake investment and attached importance to academic achievement. It could be inferred that a full degree study abroad is more likely to yield high receptive L2 contact among international students, compared with exchange programs.

In the same vein, students reported high academic L2 contact in the interview, rather than leisure L2 contact, in stark contrast with many studies which seldom mentioned academic workload (Conroy, 2018). Except for writing homework, their academic contact encompassed a wide range of tasks including listening and reading in English. It demonstrates that many participants rely on external pressure to use English in their free time, instead of using it as a voluntary act. When the pressure is alleviated, their English using frequency will decrease. Therefore, it could also be inferred that their general L2 use would see ebbs and flows between vacations and term time.

There are three possible reasons for the high academic L2 contact in the findings. Above all, the program feature influences the type of L2 contact. The sample in the present study had genuine and continuous pressure of fulfilling academic requirements to be enrolled in a university for formal degree study, while participants in other SA studies already secured the university for enrolment before going abroad. The subject courses of the current sample are reading-intensive and lecture-based. As a result, these learners spent more time doing homework to secure an unconditional offer, which meant less time for L2 socializing. Secondly, due to the change in the geopolitical landscape since the Trump administration, many Chinese students could not study in the US and applied to British universities instead, making the application competition fiercer than ever since 2019 (Lou, 2021). Thus, the participants felt more stress in obtaining high grades to gain a competitive edge in the application. Lastly, this result revealed the low integration level of the current sample and they rarely had meaningful English speaking and listening opportunities out of class, except for service encounters which were

ranked the second most frequent L2 activity by them. Therefore, this result did not necessarily suggest that the participants were extraordinarily hardworking and diligent. It only implied that their other forms of English contact were even less than their academic English contact.

7.2.1.1b Developmental profile

Regarding developmental profile, quantitative data showed that the informal L2 use did not differ significantly from T1 to T2, in line with the findings in Wright and Schartner (2013). Notably, there was a significant increase in their L1 social media use and a significant decrease in L2 speaking contact with international friends. This decline in language contact in SA literature has been commonly established. García-Amaya (2017) identified a similar trajectory of language contact change in a six-week study abroad program: American students used their L2 the most in the first half of their program and then reduced their L2 use in the last two weeks. So did N. Taguchi (2008) who recorded a decrease in L2 use in speaking, listening, and writing after 11 weeks of studying abroad. In the present study, it was speculated that at T2, the excitement of meeting multicultural friends in a new environment wore off over time and the cultural barriers emerged. “L2 learners who did not share common interests with their classmates may have struggled to develop new topics of conversation beyond the initial acquaintance” (García-Amaya, 2017, p. 69). Another possibility was that due to the COVID-19 outbreak in China, many Chinese participants shifted their attention from building connections in the UK to following the news and caring for their families in their home country, trading off the time and efforts invested in international friendship.

When it comes to individual trajectories, the interview showed consistent socialization patterns of each individual and diverging lifestyles between participants from the start of the program. The majority of interviewees naturally formed segregation and stuck in the comfort zone of co-nationals. Many Chinese students abandoned their cultural goals of SA and focused on academic endeavors (Mikal et al., 2015). In contrast, the high-social-contact minority embraced multiculturalism and enjoyed the cross-cultural

friendship. Another notable divergence at T2 is the attitude toward coursework. While many interviewees still took coursework seriously and hope to excel, two or three unmotivated participants like AC8 (and possibly other participants who did not take the interview) just wanted to spend minimal effort to pass the exams. This result echoes the L2 motivation decline found in other ERASMUS students in the UK (Hessel, 2016), which demonstrates the common difficulty of maintaining strong L2 motivation.

7.2.1.2 Social networks

7.2.1.2a Social network patterns

The findings from this study largely corroborate existing research on social networks, showing the common challenges faced by international students in an Anglophone country, especially Chinese students. The participants mainly had co-national friends, followed by international friends and rarely British friends, which supports the concentric circles model by Coleman (2013) and other literature (Gareis et al., 2011). The functional model by Bochner et al. (1977) which proposed that SA learners had more local friends than international friends was refuted by the current study.

Regarding each category of friends, Chinese students are known for strong co-national networks during SA (Gao, 2017; Wright & Scharfner, 2013). Above all, the logic behind the formation of the social networks of these students is *homophily* (McPherson et al., 2001). The benefits of maintaining L1-speaking social networks during SA were well documented in the literature, such as the provision of emotional and informational support, alleviating homesickness, social companionship (Maundeni, 2001), reducing uncertainty, and creating a sense of belonging (Rienties & Nolan, 2014). However, in the long term, the L1 peers would hinder successful cross-cultural adaptation and L2 interaction of the learners (Sauer & Ellis, 2019). In line with existing literature, the current sample did not make much progress in making L2-speaking friends and maintained primarily L1-speaking cliques over the first two terms of studying abroad (McManus et al., 2014; Rienties & Nolan, 2014).

In terms of international friends, the superficial and ephemeral relationship of the participants was usually limited to campus (McManus et al., 2014). As they are just ‘weak ties’, they tend to rely on face-to-face contact for communication, which means they only talk to each other when they meet (Van Cleemput, 2010). In the current study, the limited L2 socialization often came from group discussion with international classmates and the drinks or meals after that, but not other forms of socialization. It seems that the habits and hobbies of the participants from different countries had so little overlap with each other that the most natural reason for them to make plans together was compulsory commitments like a group project. In contrast, some SA populations were found to be bonded with international friends in the host university and formed supportive relations (Mitchell, 2015; Schartner, 2015; Xie & Liu, 2021). The commonality of those participants was their high L2 proficiency (and possibly outstanding communicative adaptability). For example, the participants in Schartner (2015) were enrolled in a MA program in Applied Linguistics with a minimum score of IELTS 6.5; participants in Xie and Liu (2021) studied in one of the top five universities in the world with a minimum score of IELTS 7.0; the sample in Mitchell (2015) were highly motivated students majoring in the target L2. Therefore, the social network patterns of the current sample might be more representative of the mass Chinese students in the UK with lower English proficiency.

In the present study, close cross-cultural friendships tended to be found among African students, Muslim students, or Confucius Asian (China, Korea, and Japan) students themselves. As the global geopolitical tension intensified, some media coverage might have deepened the stereotypes or negative depiction of certain countries, making the ‘ice’ harder to ‘break’. For example, Participant B11 had a close-knit international friends’ circle, not because of her proactive attitude or cross-cultural awareness, but still because of the *homophily* with her friends. There were very few students from Ivory Coast in the UK, so she formed a clique with friends from Nigeria, Zambia, Uganda, and Angola. She had a poor relationship with her Chinese flatmates and they barely talked. This finding revealed that in certain circumstances, SA participants could enjoy

the dual benefit of opportunities of using L2 and shared cultural values with friends. It can be inferred that the lack of co-national friends abroad might be a catalyst in ‘forcing’ learners to socialize in English with students from countries with similar cultures. The phenomenon of ELFSA (English as a Lingua Franca during Study Abroad) was observed by Köylü (2021) which highlighted the linguistic benefit of speaking English with non-native English speakers, which is supported by the current study.

Lastly, local British people were almost beyond the reach of the sample who were newcomers and peripheral participants in the host country. Previous research showed that 56% of international students in London (Furnham & Alibhai, 1985) and 70% of international students in Oxford (Bochner et al., 1985) did not have local friends after staying there for one year. Among 29 Japanese students who studied in New Zealand for three months, only two of them made local friends (Tanaka, 2007). Actually, even if there were local students in the same program, it was unlikely that they would mix and mingle with international students. In a large-scale study in a university in southern England, Rienties and Nolan (2014) found that the segregation between Confucius Asian students and local students was clearly visible.

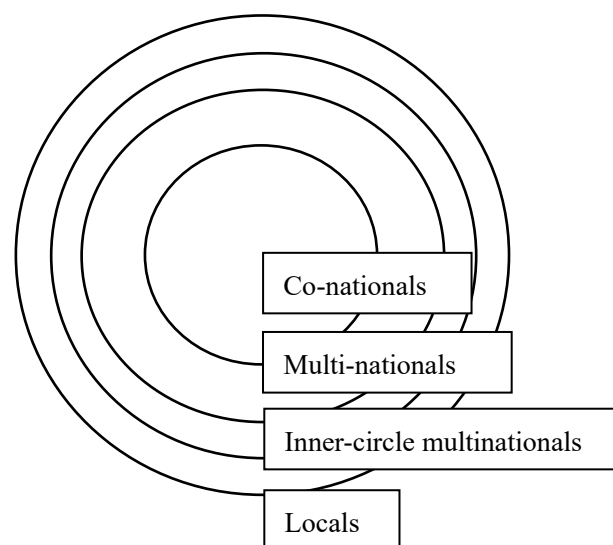
This phenomenon could also be explained by the *homophily* theory that local students have very little in common with the Chinese students (Spencer-Oatey & Xiong, 2006). For example, Chinese interviewees in the present study generally showed little interest in the British party and pub culture, which made it hard to make local friends (Gao, 2017; Schartner, 2015). Only those who formed romantic relationships with local people seemed to be exceptions to forming deep relationships (Kinger & Blattner, 2009). Lastly, the current sample from peripheral countries had another deficiency in attracting local friends compared with Anglophone students abroad (Dewey, Belnap, et al., 2013; Mitchell, 2015), because their L1s were not a Lingua Franca that some local people might desire to practice. American students studying in the UK also have a linguistic advantage in making local friends (Furnham & Alibhai, 1985). On the contrary, Chinese students often face impotence, frustration, and anxiety in speaking with local people in the UK (L. Brown, 2008). Therefore, more psychological

adjustments may have been required from these students to better integrate into the target community.

7.2.1.2b Social network model

Taking a step forward from Coleman (2013) (Figure 2.10), this study hypothesizes a social network model for non-native English-speaking students in the UK (see Figure 7.1). Its difference from Coleman (2013) is that it adds one more circle of friends by dividing multi-nationals into native English speakers (Outer Circle and Inner Circle of English) and non-native English speakers (Expanding Circle of English). It was originally hypothesized that the socialization barrier would lie between the Expanding Circle and the Outer Circle of English, while the data suggests that the main watershed of friendship establishment was between the Outer Circle and Inner Circle of English.

Figure 7.1 The social networks model for NNS students in the UK



Quantitative data in Section 5.1.2.3 revealed that students had the most friends from the Expanding Circle of English, some friends from the Outer Circle of English, and the least friends from the Inner Circle of English. This finding supported the original hypothesis that non-native English speakers from the Expanding Circle of English were more likely to befriend each other. Between Outer Circle friends and Inner Circle friends other than local British, it seems that for the non-native English-speaking students, friends from the Outer Circle were easier to approach. The foundation

program is a pathway for international students whose previous education qualifications are not recognized in the UK, so a low percentage of students have Anglophone education backgrounds. In qualitative interviews, it was not uncommon to see Chinese students with good relationships with friends from Pakistan or Malaysia and Participant B11 formed a close-knit clique with Nigerian and Zambian friends. On the contrary, none of the interviewees mentioned having a close Inner Circle non-British friend in the UK and two mentioned unpleasant relationships with American flatmates. In Section 5.2.3.2, the American friends that the interviewees mentioned were made before they studied in the UK. Therefore, it is speculated that although making Inner Circle friends is desirable for non-native English-speaking students, in reality, they formed close relationships with Outer Circle friends more easily than with Inner Circle friends. More quantitative data is needed to confirm this nuance in future social network studies.

Other literature on the friendship patterns of Inner Circle students abroad attests to the model from another perspective. North American university students in London nominated a large number of European and British people as friends due to cultural similarities (Furnham & Alibhai, 1985), and American students in Europe made more local friends than international friends (Savicki, 2010). Except for the *homophily* principle, the difficulty for international students in the UK to befriend students from the Inner Circle can probably be attributed to the Center-Periphery Theory (Friedmann, 1966). Since international students from the foundation program were often in a peripheral position in relation to language, culture, and values, establishing friendships with Anglophone students might require extra effort. Inner Circle students might have less interest in making international friends if they do not demonstrate near-native English proficiency for smooth communication or some attractive social capital that interests them.

7.2.1.2c Developmental profile

In terms of developmental trajectory from T1 to T2, in the current study, fossilization

was much more common than development in English-speaking social networks. This result is in line with literature that did not identify significant changes in the proportion of interaction with co-nationals and host-nationals (Diniz & Noels, 2021; Mitchell et al., 2017) but contradicts another study that observed significant development in local friendship (Hendrickson, 2018). In the present study, except for common reasons of cultural barriers and linguistic barriers, the fossilization could also be traced back to the limited time spent with L2-speaking friends due to other commitments such as study. Dewey, Ring, et al. (2013) found that enough time spent together facilitated good L2 relationships while infrequent meetups hindered friendship development. Qualitative data demonstrated that the few enhancements of L2 friendship in the current study both had some sort of common goals, such as working in a Chinese supermarket (Participant BC7) and doing a group project together (Participant AC19). These preconditions created opportunities for regular meetups and in turn, led to deeper mutual understanding.

7.2.2 Influential factors

RQ2a. What are the personal background factors that influence the language contact of the participants?

In the present study, factors influencing language contact range from nationality, overall English proficiency, motivation, to living arrangements. While it is rational that more proficient, more motivated students and those living with international flatmates have more L2 contact, the nationality difference is worth noticing. Previous SA research has usually been conducted on students from the same country and little comparison has been made between participants of different nationalities (Dewey et al., 2014; Mitchell et al., 2017). In the present study, a clear disparity emerged between the mixed-nationality sample and the Chinese students, who used English significantly less than their non-Chinese counterparts and had fewer English-speaking friends. The result supports the finding of Rienties and Nolan (2014) that the Chinese nationality of the participants could significantly predict the pattern of friendship networks in regression

analysis.

The reason might be multifold. First, the sheer number of Chinese students in the program enabled the Chinese participants to find like-minded friends from their compatriots more easily. Second, due to the *Great Firewall* in China, students in China do not have access to Facebook, Google, YouTube, WhatsApp, Instagram, Spotify, or Netflix, without a VPN. The Chinese market has fostered some equivalent entertainment platforms and Internet giants, which fundamentally shaped user preferences (Statista, 2022b). Participants from other Asian, Arabic, or African countries, on the contrary, use the same platforms as the Western people in their home country but with local content (StatCounter, 2022; Statista, 2022a). Therefore, they did not have to go through the process of switching to and getting used to new platforms for entertainment and news. Culture-wise, the origins of students in the current sample all had distinct cultures from the UK. Therefore, differences in L2 contact cannot readily be attributed to cultural shock and the challenges of settling in a new environment, as these might be comparable across nationalities.

7.2.3 The impact of COVID-19

RQ2b: How did the COVID-19 lockdown in the UK in Spring 2020 affect students' informal language contact and SA experience?

7.2.3.1 Language contact

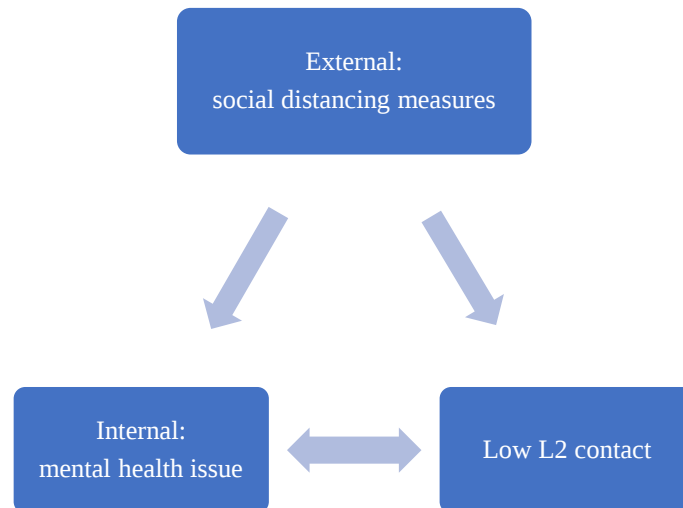
As Coleman (2013, p. 17) contends in his influential piece, SA researchers should “research whole people and whole lives” and recognize the individual and contextual variation in the SA experiences. The present study responds to this call by truthfully recording the impact of the COVID-19 pandemic on students. It is beyond doubt that this cohort of learners underwent an unconventional SA period with unprecedented isolation and fear of uncertainty. Thus, it is argued that the importance of language learning and social integration was minimized by the pandemic, and mental and physical wellbeing was prioritized by the participants during this period. Inadvertently,

this study filled a gap in SLA literature by capturing language contact and social networks of SA participants in the wake of a disaster.

Above all, environmental factors had a fundamental impact on the amount and types of language contact the learners received. Neither teachers nor students of this generation had experienced a global crisis of this scale. Therefore, the L2 language contact pattern of the sample during the pandemic was the outcome of the interplay between their living conditions, life priorities, personal idiosyncrasies, and mental status, rather than solely their intention to improve L2 or their habits of using L2. For example, some nuanced changes in the LCQ at T2 were speculated to be caused by the COVID-19 outbreak in China. The LCQ revealed an increase in L1 social media use and a decrease in time spent speaking with international friends in the LCQ. The surge of L1 use halfway through the SA program is worth discussion, as L1 speaking time with friends often peaked at the end of the sojourn abroad in preparation for returning home (García-Amaya, 2017; McManus, 2019).

Since the lockdown of Wuhan in January, Chinese students overseas were preoccupied with their families and friends in China through WeChat and other social media (e.g., Weibo). This result echoes the high proportion (71%) of respondents who had worried about their loved ones in a survey administered to international students in another British university (Schartner & Wang, 2022). Many overseas students volunteered to purchase masks and other personal protection equipment locally and shipped them back to China (CGTN, 2020), including Participant AC4 in the present study. Since their attention was concentrated on the situation in China, less time was devoted to maintaining L2 friendships. Moreover, time spent on communication in L1 on WeChat rocketed. Because Chinese students took up a large proportion of the sample in the current study, their behavioral change directly influenced the statistical differences found in the LCQ. This example further sheds light on the necessity of researching whole people and whole lives and taking contextual factors into consideration.

Figure 7.2 Reasons for low L2 contact during the lockdown



Since the lockdown, as illustrated in Figure 7.2, the social distancing measures and the consequent mental health issues resulted in the lower L2 use of the sample. The low language contact in turn worsened the depression and anxiety. Quantitative results revealed a stronger L1 presence, higher receptive L2 contact (e.g., watching movies), and lower interactive L2 contact (speaking and listening). Among the changes, the loss of speaking opportunities in English was deemed most frustrating, which undermined the value of SA. The comments from the interviewees resonate with a feeling of other international students who felt having missed out on key aspects of the SA experience (Schartner & Wang, 2022). Many participants in the current study felt unlucky being caught by this pandemic or regretful for not returning to their home country before the lockdown, which aggravated their mental issues.

As shown in qualitative findings, the most common interaction with international friends before the COVID outbreak was brief communication on campus. Because of their weak established relationship and relatively low English proficiency, calling their international friends online was not a practical alternative during lockdown to maintain L2 speaking contact. While SA learners usually regret not having more contact with L2 speakers after they finished studying abroad (Ife, 2000), in the present study, the participants probably felt more powerless and disappointed about the situation because their normal study halted when they were still abroad. The L2 interaction they had and the fluency they achieved for the first two terms almost represented the best they would

have for the whole academic year.

Another big loss of the lockdown was no chance to speak with service personnel. In previous literature, the value of conversations in service encounters was deemed limited because the information exchange was superficial (Dörnyei et al., 2004). However, as shown in Section 5.1.3.2c, this aspect of language contact was valued by some participants and every little success in the negotiation of meaning in such contexts increased their confidence. Wright and Schartner (2013) also found that learners had pleasure in their improvements in pragmatic or contextual understanding of everyday speech in predictable contexts and growing familiarity with what to say. During the lockdown, students lost this benchmark to check their English development in real life, which made their English oral fluency rustier and made them less confident in speaking English.

When it comes to receptive L2 contact, the increased quantity did not necessarily lead to more gains, because the participants were not paying attention. Being stuck in their room all day long, many students had sleeping disorders (Schartner & Wang, 2022) and they were stressed about maintaining physical health. They killed the time by watching TV shows or movies, or playing games. Some students let TV series play just so they could hear people talking and feel less lonely. Briggs Baffoe-Djan and Zhou (2020, p. 81) argue that “the pillars of quality of language contact in a SA setting are *diversity*, *significance*, and *activation*”. According to this operationalization, although the multimedia input involved diverse topics with regular activation, it was not socially or emotionally significant for participants in the lockdown. Therefore, even if they received abundant multimedia L2 contact, their attention was not focused and the benefit of this exposure might be negligible. On a positive note, some interviewees mentioned that their frequency of browsing COVID-19 news saw a rise. Based on usage-based approach, the repetitive exposure of COVID-related PVs (e.g., *roll out*, *rule out*, *run out*) might facilitate their retention of certain PVs.

When it comes to academic activities, Emergent Remote Teaching became a global

phenomenon during the pandemic, and this also impacted their exposure to the L2 in academic settings. As reported in other education literature since the COVID-19 (W. Stewart & Lowenthal, 2021), participants in the current study found it less interactive and harder to focus on online classes and expressed nostalgia for in-class teaching. With this negative attitude and practical technical constraints, it was hard for them to absorb much useful English input from the online classes. What stood out from the interview excerpts was the pragmatic orientation of some Chinese participants that their commitment to study was directly influenced by their university application outcome (see Section 5.2.3.1).

7.2.3.2 Social networks

The social networks of the participants during the pandemic were quite delicate. A few changes are consistent with findings on Swiss university students in the lockdown (Elmer et al., 2020). Firstly, students were more socially isolated, which means their interpersonal contact frequency in both L1 and L2 saw a decrease. Secondly, students had smaller social networks both in L1 and L2. The large group in-person socialization before the COVID-19 was impossible and the participants barely contacted friends that they were not close to. Finally, strong ties were strengthened and maintained during the lockdown. “As face-to-face interactions and random encounters are minimized due to the social distancing measures, it is likely that individuals focus on those relationships that are spatially close, most meaningful, or most established” (Elmer et al., 2020, p. 3). It should be noted that about 80% of the participants in Elmer et al. (2020) were Swiss nationals who were familiar with the domestic environment without language barriers. They still experienced high levels of depression, anxiety, stress, and loneliness due to the COVID-19. Thus, it could be speculated that the mental health issue was worse for the international students in the present study.

Social support theory will be mainly drawn upon in this section to explain the ebbs and flows of the social network changes at T3. Social support can be categorized into four types, ranging from emotional support, instrumental support (i.e., tangible aid and

services), informational support, and appraisal support (Lakey & Cohen, 2000). According to the literature, effective social support is likely to be provided by people who are socially similar to the support recipients and who have experienced similar stressors or situations (Thoits, 1995). Because the support provider had a good empathic understanding of the support recipient, the support they provided is more likely to be in concert with the needs and values of the recipient.

The proposed types of social relationship changes in Table 5.20 echoed these principles in social support theory. To be specific, close physical proximity and high intensity of previous relationships (Walker et al., 1993) enabled the support provider to understand their friends better and provide information, assistance, encouragement, and companionship during the lockdown. The families who were overseas provided students emotional support (e.g., empathy, love, trust, and caring) but no instrumental support. Thus, their contact frequency was also higher than the pre-COVID time, but not the highest. On the contrary, those classmates who had flown back to their home countries and were not close to the participants could not provide any informational support or instrumental support for the participants. Therefore, their friendship naturally faded away. Lastly, the relationship with those who were still in the UK but were not close to the participants remained the same. They could occasionally exchange information, but they were not in the position of providing emotional support or appraisal support which required more understanding of each other. Two points are worth attention here. First, during the lockdown, language learning was not prioritized by learners. Therefore, those who maintained close English-speaking friends were those who already established deep trust before the pandemic (e.g., B11, AC12). They spoke English just for communication, not for the sake of practice. Second, since all the participants were going through this crisis for the first time, their support was often mutual. Each student was both a giver and a taker at the same time.

7.2.3.3 SA experience

Qualitative data lend support to the emerging research that documents the link between

the pandemic and heightened anti-Asian racism (Elias et al., 2021). Rising nationalism and populism have been spreading worldwide over the last few years and thrived during times of crisis (Elias et al., 2021), which was also manifested in media coverage of racial slurs, hate speech, physical attacks (BBC, 2020), and stigmatization of China by a few political leaders. The first wave of the pandemic exacerbated racism and xenophobia toward minority groups in the UK. Since Chinese students are the main target of the negative incidents, this section will focus on their SA experience during the lockdown.

At T3, one fifth of the Chinese interviewees shared discrimination incidents, which corroborated another study that one out of four international students in the UK has experienced COVID-related discrimination (Schartner & Wang, 2022). In public spaces, the unpleasant experiences with strangers can be viewed from the perspectives of general racism and neo-racism. Before the national-wide coronavirus outbreak, the discrimination incidents shared by the interviewees were based on general racism that targets their Chinese race. Since the lockdown, the examples mentioned by interviewees can be explained by neo-racism. *Neo-racism*, also called cultural racism, is “racism without race” (Balibar & Wallerstein, 1991). It concerns racist actions against those who do not conform to the local habits, norms, and traditions. During the lockdown, some Chinese students were extraordinarily cautious and wore raincoats or goggles outdoors to prevent contracting the virus (Section 5.2.3.3). Confronted with the unfriendly reactions to their behavior, Participant AC14 tried to comply with local norms by keeping her outfit visually acceptable while Participant AC4 resisted abiding by the local expectations and kept wearing goggles. Their different reactions marked their opposite attitudes toward sociocultural integration.

With rising nationalism in China, overseas Chinese students also had patriotism and pride (Jiang, 2021). It is argued that an environment of growing anti-Chinese sentiment potentially had some subtle impact on the integrative motivation and long-term life planning of Chinese students. To be specific, more students would consider returning to their home country to work and live after completing their studies, where they could

have more equal social status and safer physical surroundings. Less effort might be invested into L2 learning and making L2 speaking friends in the future if they made this decision. Based on personal communication with the interviewees, Participant AC14 is trying to settle down in the UK after completing the master's program in London, whereas Participant AC4 returned to China and completed her master's degree entirely online. These examples can be seen as a negotiation process between the social expectation and their personal identities. Being overseas, the Chinese students had developed a deeper understanding of the cultural traditions and societal differences between China and the UK, and became clearer about the pros and cons of different cultures. In hindsight, this crisis accelerated their discovery process of which lifestyle they would fit in more.

7.3 The relationship between phrasal verb and language contact

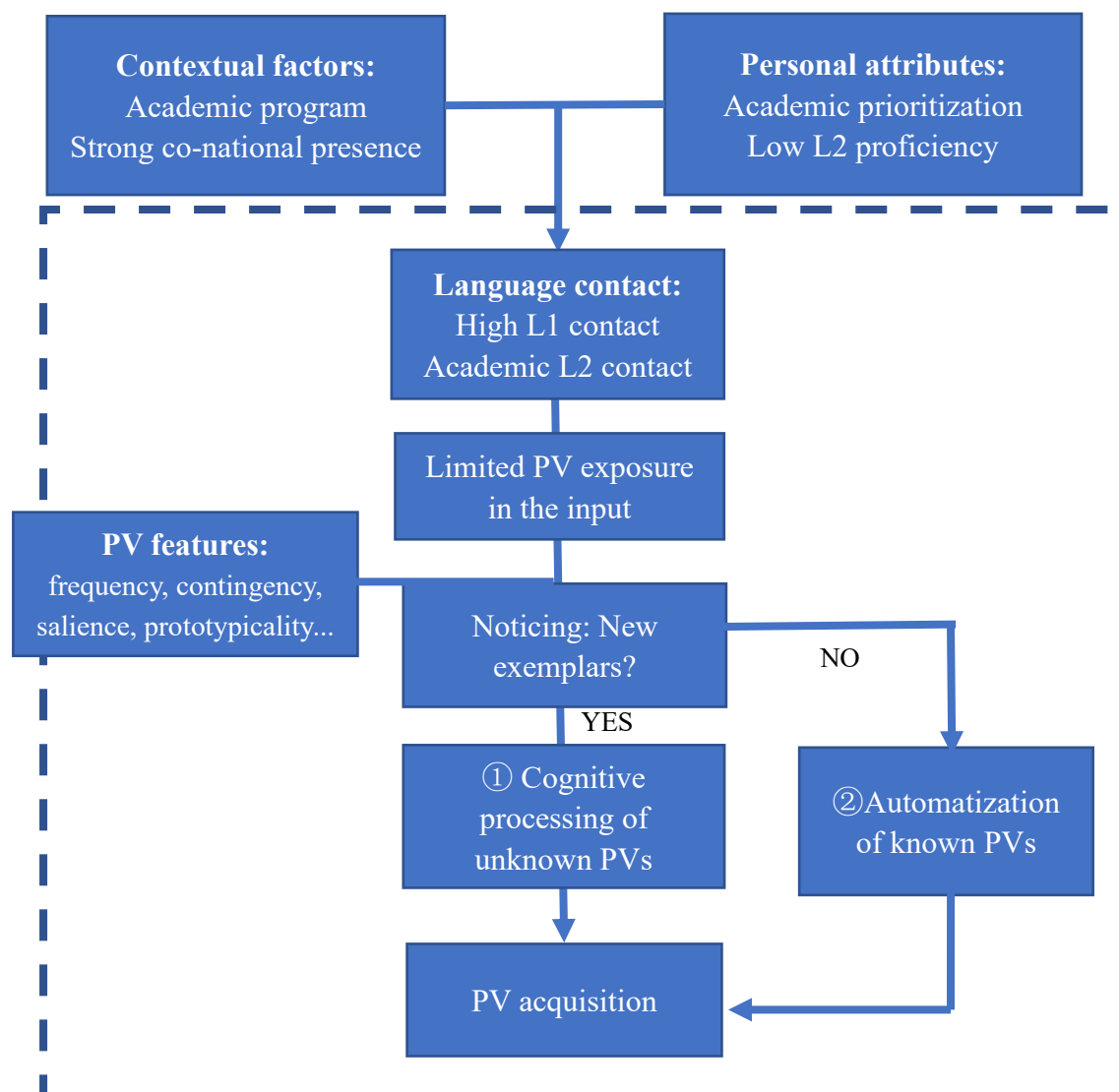
RQ3: What linguistic factors (corpus frequency, semantic opacity), personal attributes (overall English proficiency, L1), and L2 exposure (language contact, social networks) influence PV development in these participants during study abroad?

The ultimate quest that guides the present study is to what extent studying in an English-speaking country and using English out of class contributes to the acquisition of informal expressions such as PVs of English L2 learners. From T1 to T2, quantitative results demonstrated a significant positive relationship between language contact and PV knowledge, but the relationship between language contact and PV gains was weak and limited (see Section 6.1.1c). Due to the disruption of the COVID-19, the PV developmental trajectory at T3 was not generalizable. This section will discuss the results in relation to the first two time points, explained by some tenets of usage-based theory. Admittedly, language acquisition in SA is influenced by a plethora of factors and had large individual variations. The present study adheres to usage-based approach because interaction and input are particularly important for the acquisition of PVs in the non-instructed context.

7.3.1 Usage-based approach to phrasal verb acquisition in SA

As outlined in Chapter Two, usage-based theory contends that “cumulative experiences with language form memories that interact with cognitive learning mechanisms and ultimately drive language acquisition” (Crossley et al., 2019, p. 724). In essence, this approach recognizes the environmental influence on the mind and holds that we learn linguistic constructions while engaging in communication (Bybee, 2010). It argues that repeated language input and interactions lead to associative learning of constructions while acknowledging the compounding effects of learner attributes and input factors (e.g., contingency, salience, and recency).

Figure 7.3 A usage-based conceptual framework of PV acquisition during SA



In terms of PV acquisition during SA, the potential advantage of the SA environment is that it might trigger more real-life scenarios that increase the frequency and diversity of input and elicit deeper interaction with the L2 for the learners. PVs permeate the daily discourse of people in the UK. However, how much of this input is available to international students is what this study intends to find out. The framework in Figure 7.3 is adapted for the current sample based on a general framework in Zhou and Briggs Baffoe-Djan (2022). This framework proposes that the incidental acquisition of PVs during SA is influenced by contextual factors, learner attributes, and the interplay between PV features and cognitive attention. Most learners recognized that they did not pay attention to PVs in particular, suggesting the PV gains that they made came from implicit learning which is an acquisition process that takes place naturally, simply, and without conscious operations (N. C. Ellis, 2002). The participants generally had high receptive English contact from academic study and less interactive English contact due to their segregated social circles. As PVs usually consist of simple words and were often separable, learners did not notice PVs due to lower salience and students' focus on the accumulation of one-word academic vocabulary. In this condition, automatization of known PVs might be more common than cognitive processing of unknown PVs.

Students' different performances on some low-frequency and high-frequency PV items lend support to usage-based theory. In the PV test, *run out* had an above-average score (mean > .50) in both receptive and productive tests, which is unproportionate to its relatively low corpus ranking (103rd in COCA and 86th in BNC) among the tested PVs. Based on the interview, *run out* had a high frequency in grocery/supermarket shopping in the daily life of the students. It also has high salience as a non-separable PV and high contingency in expressing the meaning of something being used up, making it easier to be acquired than other abstract PVs. Two interviewees commented that they were familiar with another meaning of *make out* (kiss and cuddle) because of movies and TVs, but not the tested meaning (manage to see or hear) which is the most frequent meaning in the native speaker corpus. On the contrary, *pin down* had the lowest score in receptive and productive tests (mean < .10), implying its absence in students' L2

input. The verb *pin* also had low prototypicality compared with other lexical verbs in forming PVs (e.g., *go*, *get*, *make*). Three interviewees mentioned that they had never seen this PV before, which explained the lack of knowledge on this item.

7.3.2 Language contact and PV development

As shown in correlation results, at T1, only the receptive contact factor had a weak relationship with PV scores, while speaking contact and writing contact did not. This result could be interpreted from two perspectives. Firstly, the overall marginal relationship between language contact and PV performance at T1 might be attributed to the fact that students were still settling down in a new environment and much of their attention was paid to miscellaneous tasks which were time-consuming but not input-rich, such as purchasing homewares and daily supplies and registering at different organizations. Students usually completed these tasks with their compatriots, involving little L2 use. Therefore, their meaningful L2 use at this stage was still low and irregular. Secondly, receptive contact contained some native-like L2 input. Since students rarely had the courage to speak English and were required to write academic essays, the amount of speaking contact and writing contact were not associated with PV performance. At T1, overall English proficiency was the only predictor of their PV performance in multiple regressions. It suggested that at T1, the individual variation in the SA environment had not exerted a measurable effect on the PV scores yet and their test performance was mainly predicated on their previous knowledge.

At T2, as students settled down and formed regular daily routines following busy course schedules, stronger statistical relationships between language contact and PV scores were detected. Previous studies reveal that frequencies of reading (Garnier & Schmitt, 2016; Peters, 2018; Schmitt & Redwood, 2011), watching TV and movies (Fernández & Schmitt, 2015; Peters, 2018; Schmitt & Redwood, 2011), and using the Internet (Peters, 2018) are positively related to vocabulary knowledge. The present study supports existing studies by identifying the positive role of reading contact (including browsing websites) and multimedia contact through correlation and regression analyses.

Multimedia contact plays a unique role in facilitating PV acquisition because movies, songs, and TV series often contain lots of spoken discourse, including PVs. On social media sites, the presence of PVs is also common, in posts of short videos, memes, and informal comments. Therefore, learners with high leisure L2 contact could be attuned to frequent PV exemplars and be familiar with the contexts of using them.

On the other hand, there might be two reasons why productive contact was not identified as a predictor of PV gain at T2. First, students maintained a profile of low English speaking contact from T1, and their conversations were often limited to small talks with classmates and superficial service encounters. The low frequency of speaking English suggests that although speaking with British people was a direct way of learning PV usage, this input was rarely available to the participants. Without a good relationship with the interlocutor, even if students occasionally produced some unnatural PVs, they were usually not corrected if they were understood. When they heard others talking, they focused on catching up with the holistic idea of the speech rather than asking about the specific meaning of phrases. In this case, the benefit of speaking English was further compromised. Secondly, students reported high writing contact, but a previous pilot study has suggested this is not beneficial for PV acquisition (S. Zhou, 2018). Students were discouraged from using PVs in writing by teachers and many non-native English-speaking students unconsciously avoided using PVs due to cross-linguistic interference (Liao & Fukuya, 2004). Therefore, writing contact did not contribute to PV acquisition.

Certain types of language contact (reading contact factor and listening to podcast/radio) at T2 were weakly related to productive PV gains of the participants, corroborating the non-existing or marginal relationship between formulaic language gain and language contact in the literature (Arvidsson, 2019; N. Taguchi et al., 2013). Interestingly, these reading and listening activities did not involve sensational animations or pictures in TV series or movies. Therefore, the participants who regularly engaged in these L2 activities relied on the language itself to obtain information, so they might pay more attention to the meanings of expressions. Coincidentally, two highly motivated

interviewees mentioned that they listened to English in their spare time. For example, in Section 6.2.2a, AC4 shared at T2 that she listened to two podcast channels regularly to learn English from adapted international news and the topics encompassed technology, economics, celebrities, films, songs, and books. AC12 shared at T3 that he listened to a podcast channel called ‘Talk Art’ hosted by a British actor he liked. Two hosts invited different guests every time and they talked about different forms of art, such as movies, TV celebrities, or painters, which he enjoyed listening to. Teng (2016) found that the context in which learners encounter words plays a significant role in their word learning. The words that are encountered in a rich and clear context with relevant cues to the word meanings are more likely to be learned.

If T3 data collection was normally conducted, a cautious assumption could be made that the correlation between language contact and PVs would be stronger due to larger individual variations. Some literature has related longer stays to higher integration levels and larger vocabulary gains. For example, British students studying in Spain for two terms (eight months) made three times larger lexical organization gains and achieved a higher level of integration than those who spent one term (four months) abroad (Ife et al., 2000). Spanish exchange students in the UK for one academic year did not witness significant gain in written lexical richness until T3 and the participants that spent less time with L1-speaking friends improved their lexical richness more than those who did (Serrano et al., 2012). Although McManus et al. (2021) hold that fluency and lexis benefit most from SA because SA settings tend to be communicatively-driven and meaning-focused, the present study showed that such benefit depends on the aspect of vocabulary knowledge being measured, the SA program features, and participant characteristics.

It should be noted that multiple regression demonstrated that productive tests had a higher model fit than the receptive tests, suggesting that productive tests had fewer confounders. As mentioned in the literature review, Schmitt & Redwood (2011) also found corpus frequency explained a larger variance (20%) in the regression model of productive PV scores than receptive PV scores (9%). Based on the caveats discussed in

Section 7.1.1, the test design and learner fatigue potentially added more confounders to the receptive test result. For example, some students might guess the right answers, while some students did not bother to attempt the items that they were not confident about, leading to less variance explained by the model.

Lastly, the mixed-effects modeling of data from two time points together revealed the intricate influence of PV features, learner attributes, and exposure intensity on PV scores. Across productive and receptive tests, three predictors were consistent: corpus frequency, the amount of language contact, and overall English proficiency. In contrast, semantic opacity, L1, and network multiplexity were not significant predictors. What stands out in this analysis is the slightly varied role of time point (i.e., lengths of stay) on receptive and productive PV scores. For productive PV scores, lengths of stay itself was a significant predictor, suggesting that students, in general, were better at recalling PVs at T2 than T1. It is possible that living abroad required students to constantly activate the storage in their lexicon to deal with all sorts of daily and academic scenarios. Therefore, they became better at recalling/retrieving the PVs that they knew. As for receptive PVs, there was an interaction effect between time point and language contact, suggesting the influence of the L2 exposure became stronger as the learner stayed in the host country longer. It is speculated that as students formed regular routines after settling down, individual variations on L2 engagement were larger. Those with more L2 exposure were more likely to have higher receptive PV knowledge.

While those with strong English-speaking cliques improved their PV unconsciously in daily conversations, it is interesting to note that some hardworking students with high academic L2 contact also made some gains in PVs. It might be partly attributed to the fact that IELTS exams have a speaking component, so students actively tried to speak English with their classmates and teachers to improve their speaking scores. Participant AC14 made friends with international classmates and improved in holding conversations with the General Practitioners, the council, and other authorities on the phone at T2. Although the direct reason for their increased English interactive contact was to increase IELTS scores, they improved different aspects of their linguistic

competence, including PVs. This finding showed that as long as the L2 assessment is well-designed and has high importance, students could be motivated to improve their well-rounded ability in L2. The Chinese students in this program were pragmatic when it comes to academic study. Therefore, including more speaking components in the course requirement might encourage those introverted students to speak English more in daily life. As reviewed in the literature review, American students studying abroad in Egypt were required to speak Arabic with locals for at least two hours per day, which significantly increased the L2 contact of the participants compared with their peers in other locations in the world (Dewey et al., 2014).

As a novel attempt, the present study also examined whether language contact of the participants has a relationship with PV oral production. PV features in the learner corpus did not see a significant relationship with language contact, suggesting either their use of PVs in the specific elicitation task or their ingrained speaking idiosyncrasies were independent of the L2 contact they received. Waibel (2007) found a positive relationship between the length of stay of SA and the quantity of PVs in a written learner corpus. However, some of the counted PVs may be used incorrectly and the quality of exposure of the sojourn was unknown. Literature suggested that it is generally difficult for L2 learners to produce phrases orally (Schmitt, Grandage, et al., 2004). For the intermediate learners in this sample, speaking was still their weakness and their use of PV in the task seemed to be contingent. It is possible that their intensity of exposure was not strong enough to influence their oral production of PVs in just a few months. It could be speculated that for some SA learners with a low social integration level, maybe upon the completion of their studies abroad their oral production of PVs would remain far from idiomatic and see little improvement.

7.3.3 Social networks and PV development

Quantitative results demonstrated a lack of strong relationships between the L2 social network variables and PV scores or PV gains. Because no existing SA literature has probed the relationship between social networks and formulaic language development,

the result from the present study will be compared with SA literature investigating social networks and general vocabulary. McManus (2019) found that oral lexical complexity gains had a positive relationship with a stronger L2 social network at the end of a one-year SA period. It is suggested that the impact of frequent social interaction on L2 development is a slow and gradual process. Building friendships can take a long time, and perhaps only meaningful interactions that lead to sustained conversations contribute to language development (McManus, 2019). The present study differed from McManus (2019) in that his participants were British undergraduates majoring in French and over 70% of them studied in France as workplace interns or teaching assistants, and less than 30% were university exchange students. The participants in professional roles might have more opportunities to build meaningful relationships with locals than those who are only students, such as in the present study.

Comparing the dependent variable being measured, studies on linguistic gains and social networks in SA often delved into the speaking competence of learners (Baten, 2020; Isabelli-García, 2006). For example, American students studying different L2s abroad improved their self-perceived L2 proficiency, measured speaking proficiency, or native-like sociolinguistic variation, given the influence of L2 social networks (Baker-Smemoe et al., 2014; Dewey, Belnap, et al., 2013; Dewey et al., 2012; Gautier & Chevrot, 2015; Kennedy Terry, 2017). McManus (2019) also measured lexical complexity gains in learners' speaking data. In contrast, social networks did not witness a significant relationship with L2 writing gains in objective measures (Strobl & Baten, 2021). It could be summarized that the significant role of social networks in L2 development was often observed in speaking competence. The present study shows that although PVs are highly colloquial, the limited PV exposure that participants received from their L2 friendships was not enough to establish a statistical relationship on this type of formulaic language in the first two terms of the foundation program.

Except for the dependent variable, another factor to be considered is the participant characteristics. The above-mentioned studies were mostly conducted on American students studying in different countries. The participants in the present study were

primarily Chinese students for whom striking up conversations with strangers was less common in their culture. Therefore, it is possible that the role of social networks on linguistic gains abroad was influenced by the home culture of the sojourners. For example, Confucius Asian students might have a higher mental threshold in naming ‘friends’ in the SASNS, while American students might feel more comfortable listing as many friends as they have met in the instruments. Although clear criteria were given by the researcher on whether an individual could be counted as English-speaking friends, many participants in the current study felt inclined to only list those they were really close to, but not some foreign acquaintances such as their classmates, as they had very limited communication. In future studies, careful decisions should be made when investigating the social networks of more reserved or introverted learner cohorts.

Another possibility is that the social network of the current sample exerted an indirect impact on PV scores by influencing the quantity and types of their language contact. Admittedly, no causal relationship can be assumed from regression analysis, but the result that SASNS variables could predict considerable variances in the LCQ scores implied that the L2 network quality of the participants was associated with their L2 use patterns. Corroborated by the interview data, the participants that had higher speaking contact might be those with strong L2 networks and they could have a slight advantage in the productive PV test.

7.4 Summary of Discussion

This study aligns with existing literature in demonstrating the large variation in the broad SA research in SLA and among participants in a single study. As Kinginger (2011) pointed out, the SA environment tends to intensify individual differences. Unlike the formal instruction environment in the home country, SA experience is full of contextual factors beyond the control of participants which might influence L2 learning and they interplay with personal attributes, which leads to various L2 developmental trajectories.

Regarding PV development, international foundation students made significant gains

in the knowledge of form and meaning of PV, but not the use of PVs between T1 and T2. They usually attached low importance to this linguistic feature during SA. Thus, the acquisition of PVs by the participants almost entirely depended on implicit learning. Regarding larger productive gain than receptive gain, it can be inferred that consolidation of partial knowledge of PVs took place more often than the acquisition of new PV items for the current sample.

As for language contact, Chinese students, in general, faced larger challenges than students from other non-East-Asian countries in using English out of class, due to their high academic L2 contact, distinct cultural backgrounds, and strong co-national presence. The pandemic made them further retreat to their L1-speaking social circles, which hindered their L2 development. This study also proposes a modified social network model for non-native English-speaking students in Anglophone countries that making friends from the Inner Circle of English seems as difficult as making local British friends.

Finally, this study ascertained a weak relationship between language contact and PV gains among international foundation program students in the UK, but not the relationship between L2 social networks and PV gains directly, which has added to the mixed findings in SA literature (Borràs & Llanes, 2019). Qualitative data showed that learners with strong L2 friendships improved PVs unconsciously. Interestingly, a comprehensive exam design such as the IELTS can push some pragmatic and motivated Chinese students to increase L2 contact and practice English speaking. Further, despite the low speaking contact in L2, other types of language contact that require attention to the L2 itself, such as reading input and multimedia aural input, were positively related to PV gain, in line with usage-based approach of SLA.

CHAPTER VIII CONCLUSIONS

The present study examined the relationship between informal L2 contact in the study abroad (SA) environment and the development of phrasal verb (PV) knowledge among SA participants based on a few gaps in the literature identified in Chapter Two. Regarding PVs, there is a dearth of studies tracking the development of PVs in a naturalistic setting over a long period of time. Further, there was no study investigating the PV development using spoken learner corpora alongside receptive and productive PV tests. Secondly, previous SA studies were often conducted on Anglophone students studying abroad. Little is known about the informal language contact and social networks of international foundation students in the UK who did not have local classmates, except for some influential works by Jackson (2008, 2010). Thirdly, there is growing interest in the role of social networks in influencing L2 linguistic gains during SA, but there is lack of empirical research on the impact of social networks on formulaic language. This chapter will summarize how the current study contributes to the body of literature on lexical studies and SA from different aspects. The limitations, implications, and recommendations for future research will also be put forward.

8.1 Key findings and contributions

Regarding PV investigation, the present study demonstrated limited knowledge and an incremental and uneven developmental profile of PVs in the intermediate foundation students in the UK. Specifically, learners had a high level of mastery of a handful of high-frequency PVs. They were not only able to recognize and recall them, but also produce them frequently in speaking, suggesting the automatization of those items. But as the corpus frequency of PVs reduces, their command also lowered, with idiosyncratic mistakes in the standardized test and sporadic PV production in speaking elicitations, suggesting only declarative knowledge or partial knowledge of certain items. In terms of developmental trajectories, the participants made the most significant gains in PV

form recall, some gains in PV form-meaning recognition, and no measurable gains in PV use in spoken tasks. Due to the interplay between contextual factors and personal attributes, the PV gains among the sample were not impressive despite studying abroad and they mostly consolidated the knowledge of known PVs. Although the benefit of SA to PVs was detected, for the learners without high L2 proficiency or strong L2 friends' circles, their improvement in formulaic language might be slower than some other aspects of L2 competence, such as speaking fluency.

Data related to informal language contact enriched our understanding of the L2 contact pattern of intermediate-level foundation program students. Firstly, a unique finding of this study is that a significant proportion of the L2 contact of the participants, especially the Chinese students, was devoted to academic activities. In public's opinion, SA is often associated with cultural experience and traveling in the host country. But the current sample had to meet certain academic and language requirements for university admissions. Therefore, much of their L2 use was devoted to studying and they could not relax mentally until they secured an unconditional offer. These findings shed light on the distinct stakes that Chinese and Anglophone participants place on their SA experiences.

Regarding social networks, the most prominent finding from the current study is the subconscious segregation between students from different nationalities in their "L1 ghettos" and the low intensity of English-speaking friendship (Coleman, 2015; Wilkinson, 1998). The difficulty of establishing local friendships was common in previous literature, but this study further revealed the difficulty of establishing international friendships. The L2 social contact of individuals witnessed a polarized and stable pattern. A minority of interviewees had dense multi-cultural social circles throughout the program, while the majority of participants had mono-cultural friends circles and their interaction with international peers was limited to one or two classmates. For Anglophone students abroad, their social networking with locals declined somewhat over time (Mitchell et al., 2017). The current sample did not see such statistical decline, as their English social networking was generally low in the first

place. Further, the availability of co-national friendship might exert an influence on the formation of English-speaking social networks. A high density of Chinese compatriots might cause ‘preemption’, which left other international classmates fewer opportunities to approach Chinese students. On the other hand, the absence of co-national peers could potentially push the non-Chinese student to form a clique with students from adjacent countries, which boosts their English speaking frequency.

The main contribution of the present study to the literature is to shed light on the relationship between language contact and PV knowledge. Except for the significant role of overall English proficiency and corpus frequency of PVs, multimedia contact and reading contact at T2 could predict PV test performance, showing its positive relationship. The benefit of productive L2 contact was not detected in the present study, suggesting a lack of meaningful spoken interaction in English and the inability of the participants to pay attention to formulaic expressions in conversations. The marginal relationship between L2 contact and PV gains might be attributed to the interplay of homogenous participant nationalities, low L2 interaction, and the difficulty of the PV tests. Further, mixed-effects modeling revealed that a longer study-abroad period itself enhanced the retrieval of PVs in the productive PV tests, while the benefits of L2 exposure on receptive PV knowledge took effect with a certain threshold of lengths of stay.

This study unexpectedly captured the influence of COVID-19 on SA participants *in situ*. Since the national lockdown in the UK started, the participants saw a significant decrease in L2 using frequency which probably led to the stagnation of their PV development. This result demonstrated the detrimental impact of the COVID-19 on the SA experience and L2 learning of the participants. It also highlighted that students prioritized their physical and mental wellbeing over their L2 learning, which was justified by the circumstances. The principles of social support theory from psychology were applied to explain the social interaction of participants during the lockdown that they only maintained contact with a few closest friends and their families.

Methodologically, the contribution of the current study could be viewed at the macro and micro levels. Above all, the present study adopted a research design that entailed tracking a large sample of SA students via various instruments over one academic year, while many comparable studies were conducted on a smaller sample over one term. Although this study was impacted by the pandemic and participant attrition, its large initial sample size still allowed the researcher to garner a general picture of the L2 use patterns and PV command of the foundation program students over a normal semester-long SA period with relative confidence. In terms of instruments, this study devised a comprehensive web-based Study Abroad Social Network Survey (SASNS) which could be adopted or adapted in future empirical studies. It also trialed the practicality of compiling a spoken learner corpus via the IELTS speaking test. In the future, other elicitation tasks could be adopted and compared with the IELTS speaking test in terms of the ability to elicit PVs orally.

8.2 Limitations

The contributions of the study as outlined in the previous section need to be contextualized within an explicit knowledge of the study's limitations. These limitations stem from issues of sampling, instruments, and procedure, to force majeure. Firstly, the sample had a large proportion of Chinese students and a small percentage of students from other countries. Although it reflected a common reality in the international student population in British higher education, as a result, students from other countries had to be grouped as a whole when performing some analysis, which might obscure some regional differences between the participants from different cultural backgrounds. Further, the interview could not be carried out in the first language of all interviewees, which might influence the validity of the qualitative data for some participants who needed to operate in their L2. Secondly, severe participant attrition at T2 could be attributed to the poor attendance rate at Center A and an arrangement at Center B to exempt high-achieving students from attending English

classes. In other words, at Center B, the participants that completed the longitudinal testing were mainly the less proficient students, which might partly explain why the PV gain was limited. Therefore, the high mobility and unstable class attendance of university-level SA participants should be given more consideration in future study planning.

As for instruments of the current study, affordance of repeated testing was a main issue for consideration. To maintain students' interest and patience in T2 and T3 data collection, following the *principle of parsimony*, the designs of quantitative measures (receptive and productive PV tests, the LCQ, the SASNS) were kept concise. There was a trade-off between the number of tested items/the pieces of collected information and the number of instruments to be administered. For example, Köylü (2016) traded off shorter and fewer linguistic data collections (a one-minute speaking task and a 15-minute writing task in pre-test and post-test) for frequent language contact data collection (every month). In the current study, the final PV tests included 20 items; the LCQ included 18 items; the maximum number of friends that could be listed in the SASNS was five. While the current research design was deemed an optimal solution for the purpose of the study and the characteristics of the students, it should be recognized that testing more PVs or polysemous PVs could possibly capture more incremental gain in the participants' PV command.

In terms of the procedure of the data collection, the researcher was only allowed to visit each class once per term, which resulted in two limitations. Firstly, extensive observations and field notes could not be taken, so the researcher always tried to attend the students' classes during the weeks of data collection and informally collect information about their studies (e.g., textbooks, assignments, class schedules) as much as possible. Secondly, students had to finish four instruments in one sitting. Although efforts were put into keeping the data collection duration within 35-40 minutes, the reality that data collection often took place at the end of the class might still have caused some *fatigue effect* on students. Further, although there was a 20-minute interval between the administration of the productive and receptive PV tests, the possibility of

some minor influence from the earlier test could not be totally excluded. Though the procedure was already optimized based on the available conditions, a more rigorous design in future studies will be always desirable, such as administering the productive and receptive PV tests on different days.

Lastly, the impact of force majeure should not be overlooked. Due to the COVID-19 lockdown, the data collected at T3 not only shrank in sample size but also changed data collection methods from in-person to online. As a result, despite the year-long tracking period, the results had to be examined in two parts (i.e., the pre-COVID period and post-COVID period) rather than continuously. In other words, the present study could not confirm whether SA students would see significant PV gains over one typical academic year (i.e., if the COVID-19 outbreak did not happen). Similar replication studies could be carried out in the future when the SA environment returned to a non-lockdown situation.

8.3 Implications

This study provides several implications for international students and SA practitioners. Vocabulary-wise, Chinese foundation program students tend to adhere to the traditional language learning route of memorizing one-word vocabulary and attach less importance to formulaic language. This study revealed that they considered formulaic language too advanced to pay attention to and they deemed PVs too hard or unnecessary to learn. However, formulaic language plays a vital role in the native-like production of L2 utterances and full comprehension of the dialogues with English native speakers. Therefore, practitioners should help students have their awareness raised of the importance to accumulate formulaic language, including PVs, in their daily L2 exposure. A difference between the EFL and SA learning environments is the availability of native-speaking English input. If students realize the usefulness of formulaic language in informal communication and their English command can fulfill the needs of their daily routines, they might use more cognitive processing when

noticing unfamiliar but useful PVs in daily life, thereby acquiring them.

With regards to language contact, some measures should be taken to address the phenomenon that foundation program students had predominantly academic L2 contact but less non-academic L2 contact, especially the Chinese participants. It is hard to change students' language using idiosyncrasies (i.e., emphasis on one-word vocabulary) overnight, so a realistic expectation should be formed by students before studying abroad. If they only study abroad for the degree, they might not improve a lot in colloquial expressions of English. Students need to be mentally and academically prepared for the challenges in academic subjects and the reading and writing workloads of the foundation program, so that they could adapt to the study in the UK better and faster. Students should also be encouraged to exploit the non-academic side of SA life to a larger extent. Therefore, it is suggested that the teachers could remind the students regularly to keep the momentum of taking advantage of the L2-speaking environment in the UK, based on effective intervention by Jackson (2018). To increase the opportunity for meaningful interaction between students from different countries, teachers could require students to find international team members in group projects where applicable. From the course leader's perspective, they could consider including or increasing speaking components in the assessment. Considering the pragmatic intention of these participants, compulsory tasks affecting their scores could be the most effective in motivating students.

The findings of this study also have pastoral implications for universities as they return to normal study in the post-COVID SA programs. The findings suggest that more efforts from higher education institutions should be put into taking care of students' mental wellbeing. This is especially important in an era of increased potential reverse globalization, where students may feel more isolated in their host countries. When students travel to host universities in the post-COVID era, workshops on positivity, proper responses toward hate crimes, peer-supporter schemes, and counseling services targeting specific ethnic minority groups could be provided, to show support to international students' needs in the UK. Course leaders in programs where many

international students are present could organize social events and career planning talks with alumni, to assist in building students' social networks and help them establish connections with peers who might offer them emotional or informational support when needed.

8.4 Recommendations for future research

The current study contributed to both formulaic language research and SA literature. Thus, the recommendation for future research will also be made from these two perspectives. Regarding PV research, this study was conducted on a sample without *the priming effect* of PVs. In other words, students did not pay particular attention to PVs during SA, during which they did not fully exploit the language contact in the target community. Future studies could employ an intervention design targeting the dependent variables. Researchers can inform students of the usefulness of formulaic language at the beginning of the SA period and remind them regularly to pay attention to PVs in daily life. Students could use digital journals or diaries to record the PVs that they acquired in the meantime. Comparing the pre-test and post-test results of PVs on the control group and intervention groups in the SA setting might yield interesting findings. Further, the receptive and productive PV test could be administered to more advanced and other L1-speaking student samples in the Anglophone context, to unravel whether the development trajectories differ or not.

In terms of language contact, future longitudinal studies could track the L2 contact of the same participants from the foundation program to formal degree study. In exploring the L2 contact pattern of foundation program students in the UK, ethnographic studies can elucidate nuanced details of the L2 use of these participants. For example, how different are the L2 use patterns of students before exams and after exams, and how do their friendship with specific classmates evolve over time. In the future, qualitative case studies on SA participants from other peripheral countries other than China can shed light on the L2 contact changes of non-Chinese students. Students from peripheral

countries, in general, receive less attention in SA literature than British or American students studying abroad. This gap should be filled gradually in the future.

Lastly, the SASNS could contribute to the exploration of the social networks in a large sample of participants in future SA studies. Though a link between social networks and PV was not directly established yet, there is still a lot of potential in social network analysis in SA research. It is widely observed that forming strong L2-speaking social networks is time-consuming and challenging. Future research could explore the relationship between social networks and other types of formulaic language (e.g., collocation, idioms) in more advanced international students, so that the academia can accumulate more empirical evidence on the linguistic benefits of the L2 friendships established by international students.

8.5 Concluding remarks

In a nutshell, this large-scale longitudinal mixed-methods study explicates that the informal L2 contact of the international foundation students in the UK was only weakly correlated with a certain aspect of PV gains. The students' deficiencies in informal L2 contact and PV knowledge might be attributed to both internal and external reasons. There are a few original findings of this study. Firstly, an extra preparation year abroad for Chinese students is mainly dedicated to academic preparation, rather than social integration. Secondly, in the foundation program, insular co-national groups were common and a dynamic international student community was not found. Overall, as a relatively advanced L2 competence, PV development of less-integrated international students abroad is likely to be incremental in a half-year tracking period, though some benefits were documented with regards to reading contact, audio L2 input, and close-knit relationships with international friends. As Dörnyei et al (2004, p. 105) argued, "their success in acquiring formulaic sequences - and a phraseological competence in general - will depend on whether they can 'beat the odds', that is, whether they can break out of the 'international ghetto' they find themselves in." In the current study,

whether students could break away from their “L1 ghetto” seems to be what made the difference in PV competence compared with their peers.

References

- Adolphs, S., & Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. In N. Schmitt, *Formulaic Sequences: Acquisition, processing and use* (pp. 107–127). John Benjamins Publishing.
- Ali, A., Ali, S., Khan, S. A., Khan, D. M., Abbas, K., Khalil, A., Manzoor, S., & Khalil, U. (2019). *Sample size issues in multilevel logistic regression models*. PLoS ONE 14 (11): 0225427. <https://doi.org/10.1371/journal.pone.0225427>
- Allen, H. W. (2010). Language-learning motivation during short-term study abroad: An Activity Theory perspective. *Foreign Language Annals*, 43(1), 27–49. <https://doi.org/10.1111/j.1944-9720.2010.01058.x>
- 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. <https://doi.org/10.1111/j.1944-9720.2003.tb02120.x>
- Alqarni, I. R. (2017). The impact of length of study abroad on collocational knowledge: The Case of Saudi students in Australia. *Advances in Language and Literary Studies*, 8(2), 237–242. <https://doi.org/10.7575/aiac.all.v.8n.2p.237>
- American Council on the Teaching of Foreign languages. (1987). *ACTFL proficiency guidelines*.
- Amuzie, G. L., & Winke, P. (2009). Changes in language learning beliefs as a result of study abroad. *System*, 37(3), 366–379. <https://doi.org/10.1016/j.system.2009.02.011>
- Arvidsson, K. (2019). Quantity of target language contact in study abroad and knowledge of multiword expressions. *Study Abroad Research in Second Language Acquisition and International Education*, 4(2), 145–167.
- Atkinson, D. (2002). Toward a sociocognitive approach to second language acquisition. *The Modern Language Journal*, 86(4), 525–545. <https://doi.org/10.1111/1540-4781.00159>
- Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Variables affecting L2 gains during study abroad. *Foreign Language Annals*, 47(3), 464–486. <https://doi.org/10.1111/flan.12093>
- Balibar, E., & Wallerstein, I. (1991). *Race, Nation, Class: Ambiguous Identities*. Verso.
- Bardovi-Harlig, K., & Bastos, M.-T. (2011). Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics*, 8(3), 347–384. <https://doi.org/10.1515/iprg.2011.017>
- Barfield, A., & Gyllstad, H. (2009). *Researching Collocations in Another Language: Multiple Interpretations*. Springer.
- Barin, A., & Wellman, B. (2011). Social network analysis: An introduction. In J. Scott & P. J. Carrington, *The Sage Handbook of Social Network Analysis* (pp. 11–25). SAGE Publications.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and*

- Language*, 68(3), <https://doi.org/10.1016/j.jml.2012.11.001>
- Baten, K. (2020). The role of social networks and intense friendships in study abroad students' L2 use and speaking development: ERASMUS learners' perspectives. *Study Abroad Research in Second Language Acquisition and International Education*, 5(1), 15–44. <https://doi.org/10.1075/sar.18009.bat>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2021). *lme4: Linear Mixed-Effects Models using "Eigen" and S4* (1.1-27.1). <https://CRAN.R-project.org/package=lme4>
- BBC. (2020, March 3). Coronavirus: Student from Singapore hurt in Oxford Street attack. *BBC News*. <https://www.bbc.com/news/uk-england-london-51722686>
- Becker, T. P. (2014). Avoidance of English phrasal verbs: Investigating the effect of proficiency, learning context, task type, and verb type. *Asian Journal of English Language Teaching*, 24, 1–33.
- Bejarano, M., Dewey, D. P., Baker-Smemoe, W., Henrichsen, L. E., & Hall, T. (2019). Adult second language learners' social network development and perceived fluency gain in an intensive English program abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 4(2), 168–192. <https://doi.org/10.1075/sar.17012.bej>
- Benson, P., Barkhuizen, G., Bodycott, P., & Brown, J. (2012). Study abroad and the development of second language identities. *Applied Linguistics Review*, 3(1), 173–197.
- BERA. (2011). *Ethical Guidelines for Educational Research*. <http://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-EthicalGuidelines-2011.pdf>
- Berman, R. A., & Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Psychology Press.
- Bhandari, P. (2021, December 8). *How to deal with missing data*. Scribbr. <https://www.scribbr.com/statistics/missing-data/>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.
- Block, D. (2003). *The Social Turn in Second Language Acquisition*. Edinburgh University Press.
- Bochner, S., Hutnik, N., & Furnham, A. (1985). The friendship patterns of overseas and host students in an Oxford student residence. *Journal of Social Psychology*, 125, 689–694. <https://doi.org/10.1080/00224545.1985.9713540>
- Bochner, S., McLeod, B. M., & Lin, A. (1977). Friendship patterns of overseas students: A functional model. *International Journal of Psychology*, 12(4), 277–294. <https://doi.org/10.1080/00207597708247396>
- Boers, F. (2019). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (1st ed., pp. 143–157). Routledge. <https://doi.org/10.4324/9780429291586-10>
- Bolinger, D. (1971). *The Phrasal Verb in English*. Harvard University Press.
- Borgatti, S. P., & Lopez-Kidwell, V. (2014). Network theory. In J. Scott & P. J.

- Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 40–54). SAGE Publications Ltd. <https://doi.org/10.4135/9781446294413>
- Borràs, J., & Llanes, À. (2019). Re-examining the impact of study abroad on L2 development: A critical overview. *The Language Learning Journal*, 1–14. <https://doi.org/10.1080/09571736.2019.1642941>
- Borràs, J., & Llanes, À. (2020). L2 reading and vocabulary development after a short Study Abroad experience. *Vigo International Journal of Applied Linguistics*, 17, 35–55. <https://doi.org/10.35869/vial.v0i17.1464>
- Borràs, J., & Llanes, À. (2021). Investigating the impact of a semester-long study abroad program on L2 reading and vocabulary development. *Study Abroad Research in Second Language Acquisition and International Education*, 6(2), 276–297.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brecht, R. D., Davidson, D. E., & Ginsberg, R. B. (1995). Predictors of foreign language gain during study abroad. In B. F. Freed, *Second Language Acquisition in a Study Abroad Context* (pp. 37–66). John Benjamins Publishing. <https://benjamins.com/catalog/sibil.9.05bre>
- Briggs Baffoe-Djan, J. G., & Zhou, S. (2020). Close encounters of the third kind: Quantity, type and quality of language contact during study abroad. In M. Howard (Ed.), *Study Abroad and the Second Language Learner: Expectations, Experiences and Development* (pp. 69–90). Bloomsbury.
- Briggs Baffoe-Djan, J. G., & Smith, S. A. (2019). Descriptive statistics in data analysis. In J. McKinley & H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (1st ed., pp. 398–414). Routledge. <https://doi.org/10.4324/9780367824471-34>
- Briggs, J. G. (2014). *A study of the relationships between informal second language contact, vocabulary-related strategic behaviour and vocabulary gain in a study abroad context*. [Doctoral dissertation]. University of Oxford.
- Briggs, J. G. (2015a). Out-of-class language contact and vocabulary gain in a study abroad context. *System*, 53, 129–140. <https://doi.org/10.1016/j.system.2015.07.007>
- Briggs, J. G. (2015b). A context-specific research tool to probe the out-of-class vocabulary-related strategies of study-abroad learners. *International Journal of Applied Linguistics*, 25(3), 291–314. <https://doi.org/10.1111/ijal.12070>
- Briggs, J. G. (2016). A mixed-methods study of vocabulary-related strategic behaviour in informal L2 contact. *Study Abroad Research in Second Language Acquisition and International Education*, 1(1), 61–87. <https://doi.org/10.1075/sar.1.1.03bri>
- Brown, J. D. (2009). Choosing the right type of rotation in PCA and EFA. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 13(3), 20–25.
- Brown, L. (2008). Language and anxiety: An ethnographic study of international postgraduate students. *Evaluation & Research in Education*, 21(2), 75–95. <https://doi.org/10.1080/09500790802152167>

- Brown, L., & Holloway, I. (2008). The adjustment journey of international postgraduate students at an English university: An ethnographic study. *Journal of Research in International Education*, 7(2), 232–249. <https://doi.org/10.1177/1475240908091306>
- Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>
- Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, 1(2), 131–151. <https://doi.org/10.1111/j.1944-9720.1967.tb00127.x>
- Castañeda, M. E., & Zirger, M. L. (2011). Making the most of the “new” study abroad: Social capital and the short-term sojourn. *Foreign Language Annals*, 44(3), 544–564. <https://doi.org/10.1111/j.1944-9720.2011.01146.x>
- CGTN. (2020, January 31). *Overseas Chinese nationals send medical supplies to Wuhan*. <https://news.cgtn.com/news/2020-01-31/Overseas-Chinese-nationals-send-medical-supplies-to-Wuhan-NHkkyexda8/index.html>
- Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, 35(1), 3–25. <https://doi.org/10.1177/0265532216676851>
- Chiswick, B. R., & Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, 26(1), 1–11.
- ClinCalc. (2019). *Sample Size Calculator*. <https://clincalc.com/stats/samplesize.aspx>
- Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education* (8th ed.). Routledge.
- Coleman, J. A. (2013). Researching whole people and whole lives. In C. Kinginger (Ed.), *Social and Cultural Aspects of Language Learning in Study Abroad* (pp. 17–44). John Benjamins Publishing Company. <https://benjamins.com/catalog/lllt.37.02col>
- Coleman, J. A. (2015). Social circles during residence abroad: What students do, and who with. In R. Mitchell, N. Tracy-Ventura, & K. McManus (Eds.), *Social Interaction, Identity and Language Learning during Residence Abroad* (pp. 33–52). EuroSLA Monographs Series, 4.
- Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, 26(2), 227–248. <https://doi.org/10.1017/S02722263104262040>
- Collentine, J. (2009). Study abroad research: Findings, implications, and future directions. In M. H. Long & C. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 218–233). Cambridge University Press. <https://doi.org/10.1002/9781444315783.ch13>
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72–89. <https://doi.org/10.1093/applin/amm022>
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45–61.

<https://doi.org/10.1017/S0267190512000074>

- Conroy, M. A. (2018). Contextual factors in second language learning in a short-term study abroad programme in Australia. *The Language Learning Journal*, 46(3), 311–328. <https://doi.org/10.1080/09571736.2015.1118643>
- Coxhead, A., Nation, I. S. P., Woods, L., & Sim, D. (2018). Group vs. Individual: How administration procedure can affect vocabulary test scores. *New Zealand Studies in Applied Linguistics*, 24(2), 24–41.
- Craig, F., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Craig, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268–294.
- Crossley, N., Bellotti, E., Edwards, G., Everett, M. G., Koskinen, J., & Tranmer, M. (2015). *Social Network Analysis for Ego-Nets*. SAGE Publications Ltd. <https://doi.org/10.4135/9781473911871>
- Crossley, S. A., Subtirelu, N., & Salsbury, T. (2013). Frequency effects or context effects in second language word learning: What predicts early lexical production? *Studies in Second Language Acquisition*, 35(4), 727–755.
- Crossley, S., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *IRAL - International Review of Applied Linguistics in Language Teaching*, 49(1), 1–26. <https://doi.org/10.1515/iral.2011.001>
- Crossley, S., Skalicky, S., Kyle, K., & Monteiro, K. (2019). Absolute frequency effects in second language lexical acquisition. *Studies in Second Language Acquisition*, 41(4), 721–744. <https://doi.org/10.1017/S0272263118000268>
- Csizér, K., & Kormos, J. (2009). Learning experiences, selves and motivated learning behaviour: A comparative analysis of structural models for Hungarian secondary and university learners of English. In Z. Dörnyei & E. Ushioda, *Motivation, Language Identity and the L2 Self* (pp. 98–119). Multilingual Matters. <https://doi.org/10.21832/9781847691293-006>
- 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(1), 157–186. <https://doi.org/10.1111/j.1944-9720.2008.tb03284.x>
- Dabbagh, A., & Janebi Enayat, M. (2019). The role of vocabulary breadth and depth in predicting second language descriptive writing performance. *The Language Learning Journal*, 47(5), 575–590. <https://doi.org/10.1080/09571736.2017.1335765>
- Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs—A case for contrastive analysis. *Studies in Second Language Acquisition*, 7(1), 73–79. <https://doi.org/10.1017/S0272263100005167>
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and Assessing Vocabulary Knowledge*. 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.
<https://doi.org/10.2307/3588191>
- DeKeyser, R. M. (2014). Research on language development during study abroad. In C. Pérez-Vidal, *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 313–326). John Benjamins Publishing.
<https://benjamins.com/catalog/aals.13.16ch13>
- Devlin, A. M. (2019). The interaction between duration of study abroad, diversity of loci of learning and sociopragmatic variation patterns: A comparative study. *Journal of Pragmatics*, 146, 121–136.
<https://doi.org/10.1016/j.pragma.2018.08.007>
- Dewaele, J.-M., & Regan, V. (2001). The use of colloquial words in advanced French interlanguage. *EUROSLA Yearbook*, 1(1), 51–67.
<https://doi.org/10.1075/eurosla.1.07dew>
- 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(2), 303–327.
<https://doi.org/10.1017/S02722263104262076>
- Dewey, D. P. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 15, 127–148.
- Dewey, D. P. (2017). Measuring social interaction during study abroad: Quantitative methods and challenges. *System*, 71, 49–59.
<https://doi.org/10.1016/j.system.2017.09.026>
- 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). Language use in six study abroad programs: An exploratory analysis of possible predictors. *Language Learning*, 64(1), 36–71.
<https://doi.org/10.1111/lang.12031>
- Dewey, D. P., Bown, J., & Eggett, D. (2012). Japanese language proficiency, social networking, and language use during study abroad: Learners' perspectives. *Canadian Modern Language Review*, 68(2), 111–137.
<https://doi.org/10.3138/cmlr.68.2.111>
- 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(2), 269–282. <https://doi.org/10.1016/j.system.2013.02.004>
- Diao, W. (2016). Peer socialization into gendered L2 mandarin practices in a study abroad context: Talk in the dorm. *Applied Linguistics*, 37(5), 599–620.
<https://doi.org/10.1093/applin/amu053>
- Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, 26(2), 249–273.
<https://doi.org/10.1017/S02722263104262052>
- Díaz-Campos, M. (2006). The effect of style in second language phonology: An

- analysis of segmental acquisition in study abroad and regular-classroom students. In C. Klee & T. Face, *Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages* (pp. 27–39). Cascadilla Proceedings Project. <https://scholarworks.iu.edu/dspace/handle/2022/25202>
- Diniz, G. L., & Noels, K. A. (2021). Approaches to studying abroad in Brazil and their effects on language learning and intercultural contact. *System*, 97, 1–15. <https://doi.org/10.1016/j.system.2020.102443>
- Dóczi, B., & Kormos, J. (2016). *Longitudinal Developments in Vocabulary Knowledge and Lexical Organization*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190210274.001.0001>
- Doff, A., Thaine, C., Puchta, H., Lewis-Jones, P., & Stranks, J. (2015). *Cambridge English Empower Upper Intermediate Student's Book*. Cambridge University Press.
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford University Press.
- Dörnyei, Z. (2009). The L2 Motivational Self System. In E. Ushioda & Z. Dörnyei (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 9–42). Multilingual Matters.
- Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 87–106). John Benjamins Publishing Company. <https://doi.org/10.1075/llt.9.06dor>
- Dörnyei, Z., & Taguchi, T. (2009). Questionnaires in second language research. In Z. Dörnyei & T. Taguchi, *Questionnaires in Second Language Research: Construction, Administration, and Processing* (pp. 1–10). Routledge.
- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163–188. <https://doi.org/10.1177/0267658309349431>
- El-Dakhs, D. A. S. (2016). The lexical knowledge and avoidance of phrasal verbs: The case of Egyptian learners of English. *International Journal of Applied Linguistics and English Literature*, 5(1), 132–144. <https://doi.org/10.7575/aiac.ijalel.v.5n.1p.132>
- Elder, C., & Wigglesworth, G. (2006). An investigation of the effectiveness and validity of planning time in part 2 of the IELTS Speaking Test. *International English Language Testing System (IELTS) Research Reports*, 6, 1–28.
- Elias, A., Ben, J., Mansouri, F., & Paradies, Y. (2021). Racism and nationalism during and beyond the COVID-19 pandemic. *Ethnic and Racial Studies*, 44(5), 783–793. <https://doi.org/10.1080/01419870.2020.1851382>
- Elliott, V. (2018). Thinking about the Coding Process in Qualitative Data Analysis. *Qualitative Report*, 23, 2850–2861. <https://doi.org/10.46743/2160-3715/2018.3560>
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18(1), 91–126.

- <https://doi.org/10.1017/S0272263100014698>
- Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1), 1–24. <https://doi.org/10.1093/applin/ami038>
- Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. van Patten & J. Williams, *Theories in Second Language Acquisition: An Introduction* (2nd ed., p. 20). Routledge.
- Ellis, R. (2003). *Task-based Language Learning and Teaching* (1st edition). Oxford University Press.
- Elmer, T., Mepham, K., & Stadtfeld, C. (2020). Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PLoS ONE*, 15(7), e0236337. <https://doi.org/10.1371/journal.pone.0236337>
- Erman, B., & Warren, B. (2000). The Idiom Principle and the Open Choice Principle. *Text*, 20, 29–62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Failasofah, F., & Alkhrisheh, H. T. D. (2018). Measuring Indonesian students' lexical diversity and lexical sophistication. *Indonesian Research Journal in Education*, 2(2), 97–107. <https://doi.org/10.22437/irje.v2i2.6098>
- Faitaki, F., & Murphy, V. A. (2019). Oral language elicitation tasks in applied linguistics research. In J. McKinley & H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (1st ed., pp. 360–369). Routledge. <https://doi.org/10.4324/9780367824471-31>
- Fernández, G. B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *ITL - International Journal of Applied Linguistics*, 166(1), 94–126. <https://doi.org/10.1075/itl.166.1.03fer>
- Fernandez, J., & Gates Tapia, A. M. (2016). An appraisal of the Language Contact Profile as a tool to research local engagement in study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 1(2), 248–276. <https://doi.org/10.1075/sar.1.2.05fer>
- Finlayson, A. (2019). *Preparing for degree study: Analysis of Access to Higher Education Diplomas and integrated foundation year courses* (OfS 2019.20). Office for Students.
- Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 116–132). Cambridge University Press. <https://doi.org/10.1017/CBO9780511667268.009>
- Fitzpatrick, T. (2012). Tracking the changes: Vocabulary acquisition in the study abroad context. *The Language Learning Journal*, 40(1), 81–98. <https://doi.org/10.1080/09571736.2012.658227>
- Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan,

- & M. Swain, *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing* (pp. 75–93). Harlow: Longman.
- Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying abroad. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary Studies in First and Second Language Acquisition: The Interface Between Theory and Application* (pp. 91–106). Palgrave Macmillan. https://doi.org/10.1057/9780230242258_6
- Foster, P., Bolibaug, C., & Kotula, A. (2014). Knowledge of nativelike selections in a L2. *Studies in Second Language Acquisition*, 36(01), 101–132. <https://doi.org/10.1017/S0272263113000624>
- Freed, B. F. (1995). *Second Language Acquisition in a Study Abroad Context*. John Benjamins Publishing.
- Freed, B. F. (1998). An overview of issues and research in language learning in a study abroad setting. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 4(2), 31–60.
- Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, 26(02), 349–356. <https://doi.org/10.1017/S027226310426209X>
- 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(2), 275–301. <https://doi.org/10.1017/S0272263104262064>
- Friedmann, J. (1966). *Regional Development Policy: A Case Study of Venezuela*. MIT Press.
- Furnham, A., & Alibhai, N. (1985). The friendship networks of foreign students: A replication and extension of the functional model. *International Journal of Psychology*, 20(3–4), 709–722. <https://doi.org/10.1080/00207598508247565>
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(1), 155–179. <https://doi.org/10.1111/lang.12225>
- Gao, F. (2017). Negotiating participation in the local English-speaking community: A case study of Chinese learners in Britain. *Cambridge Journal of China Studies*, 12(3), 45–63.
- García-Amaya, L. (2017). Detailing L1 and L2 use in study-abroad research: Data from the daily linguistic questionnaire. *System*, 71, 60–72. <https://doi.org/10.1016/j.system.2017.09.023>
- García-Amaya, L. (2021). Exploring the connection between language use and oral performance during study abroad: Results from the Daily Language Questionnaire 2. *Foreign Language Annals*, 55(1), 1–24. <https://doi.org/10.1111/flan.12587>
- Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, 41(2), 339–359. <https://doi.org/10.1002/j.1545-7249.2007.tb00062.x>

- Gardner, D., & Davies, M. (2018). Sorting them all out: Exploring the separable phrasal verbs of English. *System*, 76, 197–209. <https://doi.org/10.1016/j.system.2018.06.009>
- Gareis, E., Merkin, R., & Goldman, J. (2011). Intercultural friendship: Linking communication variables and friendship success. *Journal of Intercultural Communication Research*, 40(2), 153–171. <https://doi.org/10.1080/17475759.2011.581034>
- Garnier, M. (2016). *English phrasal verbs: Usage, knowledge, acquisition* [Doctoral dissertation]. University of Nottingham.
- Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, 19(6), 645–666. <https://doi.org/10.1177/1362168814559798>
- Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, 59, 29–44. <https://doi.org/10.1016/j.system.2016.04.004>
- Gautier, R. (2019). Understanding socialisation and integration through social network analysis: American and Chinese students during a stay abroad. In M. Howard, *Study Abroad, Second Language Acquisition and Interculturality* (pp. 207–236). Multilingual Matters.
- Gautier, R., & Chevrot, J.-P. (2015). Social networks and acquisition of sociolinguistic variation in a study abroad context: A preliminary study. In R. Mitchell, N. Tracy-Ventura, & K. McManus, *Social Interaction, Identity and Language Learning during Residence Abroad* (pp. 169–184). EUROSLA.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *The Modern Language Journal*, 99(1), 137–157. <https://doi.org/10.1111/modl.12190>
- Gyllstad, H. (2005). Words that go together well: Developing test formats for measuring learner knowledge of English collocations. *The Department of English in Lund: Working Papers in Linguistics*, 5, 1–31.
- Hacking, J. F., & Tschirner, E. (2017). The contribution of vocabulary knowledge to reading proficiency: The case of college Russian. *Foreign Language Annals*, 50(3), 500–518. <https://doi.org/10.1111/flan.12282>
- Haider, A. S., Saed, H. A., Hussein, R. F., Al-Abbas, L. S., & Meqdadi, S. R. (2020). The impact of English proficiency on university students' use of one-word or phrasal verbs. *International Journal of Innovation*, 14(5), 1184–1196.
- Hardison, D. M. (2014). Changes in second-language learners' oral skills and socio-affective profiles following study abroad: A mixed-methods approach. *Canadian Modern Language Review*, 70(4), 415–444. <https://doi.org/10.3138/cmlr.2202>
- Hendrickson, B. (2018). Intercultural connectors: Explaining the influence of extra-curricular activities and tutor programs on international student friendship

- network development. *International Journal of Intercultural Relations*, 63, 1–16. <https://doi.org/10.1016/j.ijintrel.2017.11.002>
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303–317. <https://doi.org/10.1017/S0272263199002089>
- Hernández, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *The Modern Language Journal*, 94(4), 600–617. <https://doi.org/10.1111/j.1540-4781.2010.01053.x>
- Hessel, G. (2016). *The impact of participation in ERASMUS study abroad in the UK on students overall English language proficiency, self-efficacy, English use anxiety and self-motivation to continue learning English: A mixed-methods investigation* [Doctoral dissertation]. University of Oxford.
- Hirakawa, M., Shibuya, M., & Endo, M. (2018). Explicit instruction, input flood or study abroad: Which helps Japanese learners of English acquire adjective ordering? *Language Teaching Research*, 23(2), 158–178. <https://doi.org/10.1177/1362168817752377>
- Hogan, B., Janulis, P., Phillips, G. L., Melville, J., Mustanski, B., Contractor, N., & Birkett, M. (2020). Assessing the stability of egocentric networks over time using the digital participant-aided sociogram tool Network Canvas. *Network Science*, 8(2), 204–222. <https://doi.org/10.1017/nws.2019.27>
- Hogan, B., Melville, J., Phillips II, G., Janulis, P., Contractor, N., Mustanski, B., & Birkett, M. (2016). *Evaluating the Paper-to-Screen Translation of Participant-Aided Sociograms with High-Risk Participants*. <http://dx.doi.org/10.1145/2858036.2858368>
- Holliday, A. (2010). Analysing qualitative data. In B. Paltridge & A. Phakiti (Eds.), *Continuum Companion to Research Methods in Applied Linguistics* (pp. 98–110). Continuum.
- Houshyar, S., & Talebinezhad, M. (2013). Study on avoidance behaviour among Persian EFL Learners: Phrasal verbs in focus. *Greener Journal of Educational Research*, 3(6), 238–248. <https://doi.org/10.15580/GJER.2013.6.061713680>
- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 349–381). Blackwell Pub.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51(3), 539–558. <https://doi.org/10.1111/0023-8333.00164>
- Ife, A. (2000). Language learning and residence abroad: How self-directed are students? *The Language Learning Journal*, 22(1), 30–37. <https://doi.org/10.1080/09571730085200221>
- Ife, A., Vives Boix, G., & Meara, P. (2000). The impact of study abroad on vocabulary development among different proficiency groups. *Spanish Applied Linguistics*, 4, 55–84.
- Irujo, S. (1993). Steering clear: Avoidance in the production of idioms. *The Journal*

- International Review of Applied Linguistics in Language Teaching*, 31(3), 205–220. <https://doi.org/10.1515/iral.1993.31.3.205>
- Isabelli-García, C. L. (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* (pp. 231–358). Multilingual Matters. <https://doi.org/10.21832/9781853598531-013>
- Isabelli-García, C. L., Bown, J., Plews, J. L., & Dewey, D. P. (2018). Language learning and study abroad. *Language Teaching*, 51(4), 439–484. <https://doi.org/10.1017/S026144481800023X>
- Isabelli-García, C. L., & Isabelli, C. A. (2020). What is study abroad? In *Researching Second Language Acquisition in the Study Abroad Learning Environment* (pp. 1–22). Palgrave Macmillan.
- Iwaniec, J. (2014). Motivation of pupils from southern Poland to learn English. *System*, 45, 67–78. <https://doi.org/10.1016/j.system.2014.05.003>
- Jackson, J. (2008). *Language, Identity and Study Abroad; Sociocultural Perspectives*. Equinox Publishing. <https://www.equinoxpub.com/home/language-identity-study-abroad/>
- Jackson, J. (2010). *Intercultural Journeys: From Study to Residence Abroad*. Palgrave MacMillan.
- Jackson, J. (2017). The personal, linguistic, and intercultural development of Chinese sojourners in an English-speaking country: The impact of language attitudes, motivation, and agency. *Study Abroad Research in Second Language Acquisition and International Education*, 2(1), 80–106. <https://doi.org/10.1075/sar.2.1.04jac>
- Jackson, J. (2018). Intervening in the intercultural learning of L2 study abroad students: From research to practice. *Language Teaching*, 51(3), 365–382. <https://doi.org/10.1017/S0261444816000392>
- Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, 99, 102521. <https://doi.org/10.1016/j.system.2021.102521>
- Jensen, J., & Howard, M. (2014). The effects of time in the development of complexity and accuracy during study abroad: A study of French and Chinese learners of English. *EUROSLA Yearbook*, 14, 31–64. <https://doi.org/10.1075/eurosla.14.02jen>
- Jiang, S. (2021). The call of the homeland: Transnational education and the rising nationalism among Chinese overseas students. *Comparative Education Review*, 65(1), 34–55. <https://doi.org/10.1086/712053>
- Jiménez, A. F. J. (2010). A comparative study on second language vocabulary development: Study abroad vs classroom settings. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 19, 105–123.
- Kachru, B. B. (1982). *The Other Tongue: English across Cultures*. University of Illinois Press.
- Kamarudin, R., Majid, F. A., Mohd Zamin, A. A., & Mat Daud, N. S. (2019). L2 learners' receptive and productive knowledge of phrasal verbs. *International Journal of*

- Education and Literacy Studies*, 7(4), 144–149.
<https://doi.org/10.7575/aiac.ijels.v.7n.4p.144>
- Karim, S., & Haq, N. (2014). An assessment of IELTS speaking test. *International Journal of Evaluation and Research in Education (IJERE)*, 3(3), 152–157.
<https://doi.org/10.11591/ijere.v3i3.6047>
- Kennedy Terry, K. M. (2017). Contact, context, and collocation. The emergence of sociostylistic variation in L2 French learners during study abroad. *Studies in Second Language Acquisition*, 39(03), 553–578.
<https://doi.org/10.1017/S0272263116000061>
- Khorshidi, H. R. (2013). *Interlanguage Pragmatic Development in Study Abroad Program*. 2(3), 12.
- Kilic, M. (2019). Vocabulary knowledge as a predictor of performance in writing and speaking: A case of Turkish EFL learners. *PASAA: Journal of Language Teaching and Learning in Thailand*, 57, 133–164.
- Kim, Y. Y. (2001). *Becoming Intercultural: An Integrative Theory of Communication and Cross-Cultural Adaptation*. SAGE.
- Kinginger, C. (2008). Language learning in study abroad: Case studies of Americans in France. *The Modern Language Journal*, 92(s1), 1–124.
<https://doi.org/10.1111/j.1540-4781.2008.00821.x>
- Kinginger, C. (2011). Enhancing language learning in study abroad. *Annual Review of Applied Linguistics*, 31, 58–73. <https://doi.org/10.1017/S0267190511000031>
- Kinginger, C., & Blattner, G. (2009). Histories of engagement and sociolinguistic awareness in study abroad: Colloquial French. In L. Ortega & H. Byrnes (Eds.), *The Longitudinal Study of Advanced L2 Capacities* (pp. 223–246). Routledge.
- Köylü, Z. (2016). *The influence of context on L2 development: The case of Turkish undergraduates at home and abroad* [Doctoral dissertation]. University of South Florida.
- Köylü, Z. (2021). The ERASMUS sojourn: Does the destination country or pre-departure proficiency impact oral proficiency gains? *The Language Learning Journal*, 0(0), 1–13. <https://doi.org/10.1080/09571736.2021.1930112>
- Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Prentice-Hall.
- Kraut, R. E. (2017). The development of L2 reading skills: A case study from an eight-week intensive English program course. *Dialogues: An Interdisciplinary Journal of English Language Teaching and Research*, 1(1).
<https://doi.org/10.30617/dialogues1.1.4>
- Kremmel, B., & Schmitt, N. (2017). Vocabulary levels test. In J. I. Lontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). John Wiley & Sons. <https://doi.org/10.1002/9781118784235.eelt0499>
- Kuiper, K., Columbus, G., & Schmitt, N. (2009). The acquisition of phrasal vocabulary. In S. Foster-Cohen (Ed.), *Language Acquisition* (pp. 216–240). Palgrave Macmillan UK. https://doi.org/10.1057/9780230240780_10
- Lakey, B., & Cohen, S. (2000). Social support theory and measurement. In *Social Support Measurement and Intervention: A Guide for Health and Social*

- Scientists* (pp. 29–52). Oxford University Press.
<https://doi.org/10.1093/med:psych/9780195126709.003.0002>
- Lantolf, J. P., & Thorne, S. L. (2007). Sociocultural Theory and Second Language Learning. In B. van Patten & J. Williams (Eds.), *Theories of second language acquisition* (pp. 197–221). Lawrence Erlbaum.
- Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26. <https://doi.org/10.1093/applin/22.1.1>
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365–391. <https://doi.org/10.1111/0023-8333.00046>
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647–672. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Leonard, K. R., & Shea, C. E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal*, 101(1), 179–193. <https://doi.org/10.1111/modl.12382>
- Li, C., Li, W., & Ren, W. (2021). Tracking the trajectories of international students' pragmatic choices in studying abroad in China: A social network perspective. *Language, Culture and Curriculum*, 34(4), 398–416. <https://doi.org/10.1080/07908318.2020.1857393>
- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18(2), 85–102. <https://doi.org/10.1016/j.jslw.2009.02.001>
- Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, 54(2), 193–226. <https://doi.org/10.1111/j.1467-9922.2004.00254.x>
- Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *TESOL Quarterly*, 45(4), 661–688. <https://doi.org/10.5054/tq.2011.247707>
- Liu, D., & Myers, D. (2018). The most-common phrasal verbs with their key meanings for spoken and academic written English: A corpus analysis. *Language Teaching Research*, 24(3), 403–424. <https://doi.org/10.1177/1362168818798384>
- Liu, S. (2018). Review on incidental vocabulary acquisition through reading. *Theory and Practice in Language Studies*, 8(6), 702–711. <https://doi.org/10.17507/tpls.0806.22>
- Llanes, À. (2011). The many faces of study abroad: An update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism*, 8(3), 189–215. <https://doi.org/10.1080/14790718.2010.550297>
- Llanes, À. (2019). Study Abroad as a Context for Learning English as an International Language: An Exploratory Study. In M. Howard (Ed.), *Study Abroad, Second Language Acquisition and Interculturality* (pp. 136–154). Multilingual Matters.

- <https://doi.org/10.21832/9781788924153-008>
- Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37(3), 353–365. <https://doi.org/10.1016/j.system.2009.03.001>
- 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. <https://doi.org/10.1111/j.1467-9922.2012.00731.x>
- Llanes, À., & Serrano, R. (2011). Length of stay and study abroad: Language gains in two versus three months abroad. *Revista Espanola de Linguistica Aplicada*, 24, 95–110.
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(2), 126–141.
- Lou, K. (2021, May 20). 216,000 Chinese students study in UK as a result of US visa restrictions. *Global Times*. <https://www.globaltimes.cn/page/202105/1224057.shtml>
- Lu, Z., & Sun, J. (2017). Presenting English polysemous phrasal verbs with two metaphor-based cognitive methods to Chinese EFL learners. *System*, 69, 153–161. <https://doi.org/10.1016/j.system.2017.07.016>
- Macis, M., & Schmitt, N. (2017). Not just ‘small potatoes’: Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, 21(3), 321–340. <https://doi.org/10.1177/1362168816645957>
- Mackey, A., & Gass, S. M. (2015). Coding. In A. Mackey, *Second Language Research: Methodology and Design* (pp. 112–149). Taylor & Francis Group. <http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=2126959>
- Maeder-Qian, J. (2017). Linguistic identity changes of Chinese international students in Germany: A pilot study. *Study Abroad Research in Second Language Acquisition and International Education*, 2(2), 240–262. <https://doi.org/10.1075/sar.15024.mae>
- Maftoon, P., & Shakouri, N. (2018). *Cognitive Approaches to Second Language Acquisition: A Symbiosis between Nature and Nurture* (pp. 9–22). Brill. https://doi.org/10.1163/9789004380882_002
- Magnan, S. S., & Back, M. (2007). Social interaction and linguistic gain during study abroad. *Foreign Language Annals*, 40(1), 43–61. <https://doi.org/10.1111/j.1944-9720.2007.tb02853.x>
- Magnan, S. S., & Lafford, B. A. (2011). Learning through immersion during study abroad. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 525–540). Routledge. <https://doi.org/10.4324/9780203808184.ch32>
- Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, 43(5), 985–1008. <https://doi.org/10.1017/S0272263121000036>
- Martínez-Arbelaiz, A., Areizaga, E., & Camps, C. (2017). An update on the study abroad experience: Language choices and social media abroad. *International Journal of Multilingualism*, 14(4), 350–365.

<https://doi.org/10.1080/14790718.2016.1197929>

- Martinsen, R. A., Baker, W., Dewey, D. P., Bown, J., & Johnson, C. (2010). Exploring diverse settings for language acquisition and use: Comparing study abroad, service learning abroad, and foreign language housing. *Applied Language Learning*, 20, 45–69.
- Maas C. J., & Hox J. J. (2005). Sufficient sample sizes for multilevel modeling, *Methodology*, 1(3), 86–92.
- Maundeni, T. (2001). The role of social networks in the adjustment of African students to British society: Students' perception. *Race Ethnicity and Education*, 4(3), 253–276. <https://doi.org/10.1080/13613320120073576>
- Maxwell, J. A. (2010). Using numbers in qualitative research. *Qualitative Inquiry*, 16(6), 475–482. <https://doi.org/10.1177/1077800410364740>
- McCormick, T. (2018). Psycholinguistic, cognitive and usage-based approaches to study abroad research. In C. Sanz & A. Morales-Front, *The Routledge Handbook of Study Abroad Research and Practice* (pp. 36–47). Routledge.
- McIntosh, C. (2006). *Oxford Phrasal Verbs Dictionary for learners of English* (2nd edition). Oxford University Press.
- McLean, S., & Kramer, B. (2015). The creation of a new vocabulary levels test. *Shiken*, 19(2), 1–11.
- McManus, K. (2019). Relationships between social networks and language development during study abroad. *Language, Culture and Curriculum*, 32(3), 270–284. <https://doi.org/10.1080/07908318.2019.1661683>
- McManus, K., Mitchell, R., & Tracy-Ventura, N. (2014). Understanding insertion and integration in a study abroad context: The case of English-speaking sojourners in France. *Revue Française de Linguistique Appliquée*, 19(2), 97–116.
- McManus, K., Mitchell, R., & Tracy-Ventura, N. (2021). A longitudinal study of advanced learners' linguistic development before, during, and after study abroad. *Applied Linguistics*, 42(1), 136–163. <https://doi.org/10.1093/applin/amaa003>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Meara, P. (1990). Some notes on the Eurocentres vocabulary tests. *Foreign Language Comprehension and Production*, 103–113.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19–30. [https://doi.org/10.1016/S0346-251X\(99\)00058-5](https://doi.org/10.1016/S0346-251X(99)00058-5)
- Mikal, J. P., Yang, J., & Lewis, A. (2015). Surfing USA: How Internet use prior to and during study abroad affects Chinese students' stress, integration, and cultural learning while in the United States. *Journal of Studies in International Education*, 19(3), 203–224. <https://doi.org/10.1177/1028315314536990>
- Milroy, J., & Milroy, L. (1978). Belfast: Change and variation in an urban vernacular. In P. Trudgill, *Sociolinguistic Patterns in British English* (pp. 19–36). Arnold.
- Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, 21(2), 339–284.

- <https://doi.org/10.1017/S0022226700010306>
- Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL - International Journal of Applied Linguistics*, 107–108, 17–34. <https://doi.org/10.1075/itl.107-108.02mil>
- Mitchell, R. (2015). The development of social relations during residence abroad. *Innovation in Language Learning and Teaching*, 9(1), 22–33. <https://doi.org/10.1080/17501229.2014.995762>
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone Students Abroad: Identity, Social Relationships, and Language Learning*. Routledge.
- Moineddin R., Matheson F.I., & Glazier R.H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC medical research methodology*, 7(1): 34. <https://doi.org/10.1186/1471-2288-7-34> PMID: 17634107
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46(4), 610–641. <https://doi.org/10.1002/tesq.34>
- Morrison, N. (2021, February 9). *U.K. Universities Face Financial Loss As Brexit Hits E.U. Student Numbers*. Forbes. <https://www.forbes.com/sites/nickmorrison/2021/02/09/uk-universities-face-financial-loss-as-brexit-hits-eu-student-numbers/>
- Morse, J. (2010). Procedures and Practice of Mixed Method Design: Maintaining Control, Rigor, and Complexity. In A. Tashakkori & C. Teddlie, *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (pp. 339–352). SAGE Publications, Inc. <https://doi.org/10.4135/9781506335193.n14>
- Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology*, 15(6), 1044–1045. <https://doi.org/10.1093/beheco/arh107>
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Newbury House Publishers.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- O’Sullivan, B., & Lu, Y. (2006). The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test. In *International English Language Testing System (IELTS) Research Reports 2006: Volume 6* (pp. 1–27). IELTS Australia and British Council.
- Pallant, J. (2020). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. Taylor & Francis Group. <http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=6215124>
- Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381–402. <https://doi.org/10.1177/1362168815618428>
- Pellicer-Sánchez, A. (2019). Examining second language vocabulary growth: Replications of Schmitt (1998) and Webb & Chang (2012). *Language Teaching*, 52(4), 512–523. <https://doi.org/10.1017/S026144481800037X>

- Pérez Vidal, C., & Juan-Garau, M. (2009). The effect of Study Abroad (SA) on written performance. *EUROSLA Yearbook*, 9, 269–295. <https://doi.org/10.1075/eurosla.9.13per>
- Pérez-Vidal, C. (2014). *Language Acquisition in Study Abroad and Formal Instruction Contexts*. John Benjamins Publishing Company.
- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, 169(1), 142–168. <https://doi.org/10.1075/itl.00010.pet>
- Pizziconi, B. (2017). Japanese vocabulary development in and beyond study abroad: The timing of the year abroad in a language degree curriculum. *The Language Learning Journal*, 45(2), 133–152. <https://doi.org/10.1080/09571736.2013.786119>
- Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, 47(4), 424–438. <https://doi.org/10.1080/09571736.2019.1638630>
- Qi, Y., & Ding, Y. (2011). Use of formulaic sequences in monologues of Chinese EFL learners. *System*, 39(2), 164–174. <https://doi.org/10.1016/j.system.2011.02.003>
- Qian, D. D., & Lin, L. H. F. (2019). The relationship between vocabulary knowledge and language proficiency. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (1st ed., pp. 66–80). Routledge. <https://doi.org/10.4324/9780429291586-5>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.
- R Development Core Team. (2021). *R: The R Project for Statistical Computing* (4.1). <https://www.r-project.org/>
- Ragini, Ch. S. (2019). Developing lexical accuracy in communication using webtools. *Language in India*, 19(5), 208–212.
- Khorshidi, H. R. (2013). Study Abroad and Interlanguage Pragmatic Development in Request and Apology Speech Acts among Iranian Learners. *English Language Teaching*, 6(5), p62. <https://doi.org/10.5539/elt.v6n5p62>
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Read, J., & Nation, P. (2006). An investigation of the lexical dimension of the IELTS Speaking Test. In *International English Language Testing System (IELTS) Research Reports 2006: Volume 6* (Vol. 6, pp. 1–25). IELTS Australia and British Council.
- Regan, V. (1998). Sociolinguistics and Language Learning in a Study Abroad Context. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 4(2), 61–90.
- Revier, R. L. (2009). Evaluating a New Test of Whole English Collocations. In A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language: Multiple Interpretations* (pp. 125–138). Palgrave Macmillan UK. https://doi.org/10.1057/9780230245327_10

- Rienties, B., & Nolan, E.-M. (2014). Understanding friendship and learning networks of international and host students using longitudinal Social Network Analysis. *International Journal of Intercultural Relations*, 41, 165–180. <https://doi.org/10.1016/j.ijintrel.2013.12.003>
- Rivers, W. P. (1998). Is being there enough? The effects of homestay placements on language gain during study abroad. *Foreign Language Annals*, 31(4), 492–500. <https://doi.org/10.1111/j.1944-9720.1998.tb00594.x>
- Ryoo, M.-L. (2013). A corpus-based study of the use of phrasal verbs in Korean EFL students' writing. *The Journal of Asia TEFL*, 10(2), 63–89.
- Sanz, C., & Morales-Front, A. (2018). Introduction: Issues in study abroad research and practice. In C. Sanz & A. Morales-Front (Eds.), *The Routledge Handbook of Study Abroad Research and Practice* (pp. 1–16). Routledge. <https://doi.org/10.4324/9781315639970>
- Sasaki, M. (2011). Effects of varying lengths of study-abroad experiences on Japanese EFL students' L2 writing ability and motivation: A longitudinal study. *TESOL Quarterly*, 45(1), 81–105. <https://doi.org/10.5054/tq.2011.240861>
- Sauer, L., & Ellis, R. (2019). The social lives of adolescent study abroad learners and their L2 development. *The Modern Language Journal*, 103(4), 739–762. <https://doi.org/10.1111/modl.12589>
- Savicki, V. (2010). An analysis of the contact types of study abroad students: The peer cohort, the host culture and the electronic presence of the home culture in relation to readiness and outcomes. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 19, 61–86.
- Schartner, A. (2015). 'You cannot talk with all of the strangers in a pub': A longitudinal case study of international postgraduate students' social ties at a British university. *Higher Education*, 69(2), 225–241. <https://doi.org/10.1007/s10734-014-9771-8>
- Schartner, A., & Wang, Y. (2022). *Exploring the Impact on Academic, Psychological and Sociocultural Adjustment*. Newcastle University.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal Study. *Language Learning*, 48(2), 281–317. <https://doi.org/10.1111/1467-9922.00042>
- Schmitt, N. (2004). *Formulaic Sequences: Acquisition, Processing and Use*. John Benjamins Publishing.
- Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Palgrave Macmillan.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. <https://doi.org/10.1111/lang.12077>
- Schmitt, N., Dörnyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences—A longitudinal study. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 55–86). John Benjamins Publishing. <https://benjamins.com/catalog/llt.9.05sch>

- Schmitt, N., Grandage, S., & Durow, A. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In *Formulaic Sequences: Acquisition, Processing and Use* (pp. 127–151). John Benjamins Publishing Company. <https://benjamins.com/catalog/llt.9.08sch>
- Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *Studies in Corpus Linguistics* (pp. 173–208). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.45.12sch>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Vocabulary levels test. *Language Testing*, 18(1), 55–88.
- Schwieter, J. W., Jackson, J., & Ferreira, A. (2021). When ‘domestic’ and ‘international’ students study abroad: Reflections on language learning, contact, and culture. *International Journal of Bilingual Education and Bilingualism*, 24(1), 124–137. <https://doi.org/10.1080/13670050.2018.1447545>
- Scott, J., & Carrington, P. J. (2011). *The SAGE Handbook of Social Network Analysis*. SAGE Publications.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173–199. <https://doi.org/10.1017/S0272263104262027>
- 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(2), 133–143. <https://doi.org/10.1016/j.system.2011.05.002>
- Serrano, R., Llanes, À., & Tragant, E. (2016). Examining L2 development in two short-term intensive programs for teenagers: Study abroad vs. “at home.” *System*, 57, 43–54. <https://doi.org/10.1016/j.system.2016.01.003>
- Serrano, R., Tragant, E., & Llanes, À. (2012). A longitudinal analysis of the effects of one year abroad. *Canadian Modern Language Review*, 68(2), 138–163. <https://doi.org/10.3138/cmlr.68.2.138>
- Serrano, R., Tragant, E., & Llanes, À. (2014). Summer English courses abroad versus “at home.” *ELT Journal*, 68(4), 397–409. <https://doi.org/10.1093/elt/ccu007>
- Shively, R. L., & Cohen, A. D. (2008). Development of Spanish requests and apologies during study abroad. *Íkala, Revista de Lenguaje y Cultura*, 13(20), 57–118.
- Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. One-word verbs. *IRAL - International Review of Applied Linguistics in Language Teaching*, 45(2), 119–139. <https://doi.org/10.1515/IRAL.2007.005>
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429–458. <https://doi.org/10.3138/cmlr.64.3.429>
- Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System*, 53, 148–160. <https://doi.org/10.1016/j.system.2015.07.003>
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2018). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer-Sánchez, *Understanding*

- Formulaic Language: A Second Language Acquisition Perspective* (pp. 1–17). Routledge.
- Smith, S. A. (2014). *The nature of multi-word vocabulary among children with English as a first or additional language and its relationship with reading comprehension*. [Doctoral dissertation]. University of Oxford.
- Smith, S. A., & Murphy, V. A. (2015). Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language. *Reading and Writing*, 28(3), 347–369. <https://doi.org/10.1007/s11145-014-9527-y>
- Spencer-Oatey, H., & Xiong, Z. (2006). Chinese students' psychological and sociocultural adjustments to Britain: An empirical study. *Language, Culture and Curriculum*, 19(1), 37–53. <https://doi.org/10.1080/07908310608668753>
- Stark, T. H., & Krosnick, J. A. (2017). GENSI: A new graphical tool to collect ego-centered network data. *Social Networks*, 48, 36–45. <https://doi.org/10.1016/j.socnet.2016.07.007>
- StatCounter Global Stats. (2022). *Social Media Stats Africa*. StatCounter. <https://gs.statcounter.com/social-media-stats/all/africa>
- Statista. (2022a). *Asia: Social media market share by platform 2022*. Statista. <https://www.statista.com/statistics/1123881/apac-social-media-market-share-by-platform/>
- Statista. (2022b). *China: MAU of leading messaging apps 2021*. Statista. <https://www.statista.com/statistics/1062449/china-leading-messaging-apps-monthly-active-users/>
- Stewart, J. A. (2010). Using e-journals to assess students' language awareness and social identity during study abroad. *Foreign Language Annals*, 43(1), 138–159. <https://doi.org/10.1111/j.1944-9720.2010.01064.x>
- Stewart, J., & White, D. A. (2011). Estimating guessing effects on the vocabulary levels test for differing degrees of word knowledge. *TESOL Quarterly*, 45(2), 370–380. <https://doi.org/10.5054/tq.2011.254523>
- Stewart, W., & Lowenthal, P. (2021). Distance education under duress: A case study of exchange students' experience with online learning during the COVID-19 pandemic in the Republic of Korea. *Journal of Research on Technology in Education*, 0(0), 1–15. <https://doi.org/10.1080/15391523.2021.1891996>
- Strobl, C., & Baten, K. (2021). Writing development during study abroad: The role of language contact and social networks. *Study Abroad Research in Second Language Acquisition and International Education*, 6(1), 124–162. <https://doi.org/10.1075/sar.19015.str>
- Strong, B., & Boers, F. (2019a). The error in trial and error: Exercises on phrasal verbs. *TESOL Quarterly*, 53(2), 289–319. <https://doi.org/10.1002/tesq.478>
- Strong, B., & Boers, F. (2019b). Weighing up exercises on phrasal verbs: Retrieval versus trial-and-error practices. *The Modern Language Journal*, modl.12579. <https://doi.org/10.1111/modl.12579>
- Study in UK. (2020). *International Student Statistics in UK 2020*. Study in UK. <https://www.studying-in-uk.org/international-student-statistics-in-uk/>

- Study International. (2018, May 2). *Which country is home to the largest international student population?* Study International. <https://www.studyinternational.com/news/country-home-largest-international-student-population/>
- Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics*, 30(1), 79–99. <https://doi.org/10.1017/S0142716408090048>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235–253). Newbury House.
- Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics* (6th ed.). Pearson.
- Taguchi, N. (2008). Cognition, language contact, and the development of pragmatic comprehension in a study-abroad context. *Language Learning*, 58(1), 33–71. <https://doi.org/10.1111/j.1467-9922.2007.00434.x>
- Taguchi, N. (2018). Contexts and pragmatics learning: Problems and opportunities of the study abroad research. *Language Teaching*, 51(01), 124–137. <https://doi.org/10.1017/S0261444815000440>
- Taguchi, N., Li, S., & Xiao, F. (2013). Production of formulaic expressions in L2 Chinese: A developmental investigation in a study abroad context. *Chinese as a Second Language Research*, 2(1), 23–58. <https://doi.org/10.1515/caslar-2013-0021>
- Taguchi, N., Xiao, F., & Li, S. (2016). Effects of intercultural competence and social contact on speech act production in a Chinese study abroad context. *The Modern Language Journal*, 100(4), 775–796. <https://doi.org/10.1111/modl.12349>
- Taguchi, T., Magid, M., & Papi, M. (2009). The L2 Motivational Self System among Japanese, Chinese and Iranian learners of English: A comparative study. In E. Ushioda & Z. Dörnyei (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 66–97). Multilingual Matters. <https://doi.org/10.21832/9781847691293-005>
- Tanaka, K. (2007). Japanese students' contact with English outside the classroom during study abroad. *New Zealand Studies in Applied Linguistics*, 13(1), 36–54.
- Tang, X., Taguchi, N., & Li, S. (2021). Social contact and speech act strategies in a Chinese study abroad context. *Study Abroad Research in Second Language Acquisition and International Education*, 6(1), 3–31.
- Tavakoli, P. (2018). L2 development in an intensive Study Abroad EAP context. *System*, 72, 62–74. <https://doi.org/10.1016/j.system.2017.10.009>
- Teng, F. (2016). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *The Language Learning Journal*, 47(2), 145–158. <https://doi.org/10.1080/09571736.2016.1244217>
- Teng, F. (2020). The effectiveness of group, pair and individual output tasks on learning phrasal verbs. *The Language Learning Journal*, 48(2), 187–200. <https://doi.org/10.1080/09571736.2017.1373841>
- Thoits, P. A. (1995). Stress, coping, and social support processes: Where are we? What next? *Journal of Health and Social Behavior*, 35, 53–79.

- Tracy-Ventura, N. (2017). Combining corpora and experimental data to investigate language learning during residence abroad: A study of lexical sophistication. *System*, 71, 35–45. <https://doi.org/10.1016/j.system.2017.09.022>
- Tracy-Ventura, N., Dewaele, J.-M., Köylü, Z., & McManus, K. (2016). Personality changes after the ‘year abroad’?: A mixed-methods study. *Study Abroad Research in Second Language Acquisition and International Education*, 1(1), 107–127. <https://doi.org/10.1075/sar.1.1.05tra>
- Trentman, E. (2013). Arabic and English during study abroad in Cairo, Egypt: Issues of access and use. *The Modern Language Journal*, 97(2), 457–473. <https://doi.org/10.1111/j.1540-4781.2013.12013.x>
- Uchihara, T., & Clenton, J. (2020). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*, 24(4), 540–556. <https://doi.org/10.1177/1362168818799371>
- Van Cleemput, K. (2010). “I’ll see you on IM, text, or call you”: A social network approach of adolescents’ use of communication media. *Bulletin of Science, Technology & Society*, 30(2), 75–85. <https://doi.org/10.1177/0270467610363143>
- Verspoor, M. H., Lowie, W., van Geert, P., van Dijk, M., & Schmid, M. (2011). How to sections. In M. H. Verspoor, K. de Bot, & W. Lowie, *Dynamic Approach to Second Language Development: Methods and techniques* (pp. 129–200). John Benjamins Publishing Company. <http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=669007>
- Vu, D. V., & Peters, E. (2021). Incidental learning of collocations from meaningful input: A longitudinal study into three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, 1–23. <https://doi.org/10.1017/S0272263121000462>
- Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students* [Doctoral dissertation]. <https://freidok.uni-freiburg.de/data/3592>
- Walker, M. E., Wasserman, S., & Wellman, B. (1993). Statistical models for social support networks. *Sociological Methods & Research*, 22(1), 71–98. <https://doi.org/10.1177/0049124193022001004>
- Wallerstein, I. (2004). *World-Systems Analysis: An Introduction*. Duke University Press.
- Wang, C. (2010). Toward a second language socialization perspective: Issues in study abroad research. *Foreign Language Annals*, 43(1), 50–63. <https://doi.org/10.1111/j.1944-9720.2010.01059.x>
- Waring, R. (1998). *Receptive and Productive Foreign Language Vocabulary Size II*. <http://www.robwaring.org/papers/various/RPII.html>
- Watson, J. R., & Wolfel, R. (2015). The intersection of language and culture in study abroad: Assessment and analysis of study abroad outcomes. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 25, 57–72.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65. <https://doi.org/10.1093/applin/aml048>
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in*

- Second Language Acquisition*, 30(1), 79–95.
<https://doi.org/10.1017/S02722263108080042>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, 168(1), 33–69.
<https://doi.org/10.1075/itl.168.1.02web>
- Wei, Y. (2021). Use of English phrasal verbs of Chinese students across proficiency levels: A corpus-based analysis. *International Journal of TESOL Studies*, 3(4), 25–41. <https://doi.org/10.46451/ijts.2021.12.03>
- White, B. J. (2012). A conceptual approach to the instruction of phrasal verbs. *The Modern Language Journal*, 96(3), 419–438. <https://doi.org/10.1111/j.1540-4781.2012.01365.x>
- Wilkinson, S. (1998). On the nature of immersion during study abroad: Some participant perspectives. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 4(1), 121–138.
- Wilkinson, S. (2002). The omnipresent classroom during summer study abroad: American students in conversation with their French hosts. *The Modern Language Journal*, 86(2), 157–173. <https://doi.org/10.1111/1540-4781.00142>
- Winke, P. (2014). Testing Hypotheses about Language Learning Using Structural Equation Modeling. *Annual Review of Applied Linguistics*, 34, 102–122.
<https://doi.org/10.1017/S0267190514000075>
- Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. <http://arxiv.org/pdf/1308.5499.pdf>
- Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Routledge.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press.
- Wray, A. (2013). Formulaic language. *Language Teaching*, 46(03), 316–334.
<https://doi.org/10.1017/S0261444813000013>
- Wright, C., & Schartner, A. (2013). ‘I can’t ... I won’t?’ International students at the threshold of social interaction. *Journal of Research in International Education*, 12(2), 113–128. <https://doi.org/10.1177/1475240913491055>
- Wulff, S. (2018). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language* (1st ed., pp. 19–37). Routledge. <https://doi.org/10.4324/9781315206615-2>
- Xie, Y. (2017). *Who do you hang out with? A mixed-methods study on how Chinese students’ social networks relate to their oral proficiency gains during study abroad* [Master dissertation]. University of Cambridge.
- Xie, Y., & Liu, Y. (2021). Who do you hang out with?: How Chinese students’ social networks relate to their perceived oral proficiency gains during study abroad experiences. *Study Abroad Research in Second Language Acquisition and International Education*, 6(1), 59–90. <https://doi.org/10.1075/sar.18014.xie>
- Yager, K. (1998). Learning Spanish in Mexico: The effect of informal contact and student attitudes on language gain. *Hispania*, 81(4), 898–913.
<https://doi.org/10.2307/345798>

- Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-Analysis. *Language Learning*, 71(2), 487–536. <https://doi.org/10.1111/lang.12444>
- You, C., & Dörnyei, Z. (2016). Language learning motivation in China: Results of a large-scale stratified survey. *Applied Linguistics*, 37(4), 495–519. <https://doi.org/10.1093/applin/amu046>
- Zahedkazemi, E. (2015). Construct validation of TOEFL-iBT (as a Conventional Test) and IELTS (as a Task-based Test) among Iranian EFL test-takers' performance on speaking modules. *Theory and Practice in Language Studies*, 5(7), 1513–1519. <https://doi.org/10.17507/tpls.0507.27>
- Zarifi, A., & Mukundan, J. (2014). Creativity and unnaturalness in the use of phrasal verbs in ESL learner language. *3L: The Southeast Asian Journal of English Language Studies*, 20(3), 51–62. <https://doi.org/10.17576/3L-2014-2003-05>
- Zaytseva, V. (2016). *Vocabulary Acquisition in Study Abroad and Formal Instruction: An Investigation on Oral and Written Lexical Development* [Doctoral dissertation]. Universitat Pompeu Fabra.
- Zaytseva, V., Pérez-Vidal, C., & Miralpeix, I. (2018). Vocabulary acquisition during study abroad: A comprehensive review of research. In C. Sanz & A. Morales-Front (Eds.), *The Routledge Handbook of Study Abroad Research and Practice* (pp. 210–224). Routledge.
- Zhang, S., & Zhang, X. (2020). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, 1362168820913998. <https://doi.org/10.1177/1362168820913998>
- Zhang, X., & Wen, J. (2019). Exploring multiple constraints on second language development of English polysemous phrasal verbs. *Applied Psycholinguistics*, 40(05), 1073–1101. <https://doi.org/10.1017/S0142716419000146>
- Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, 56, 40–53. <https://doi.org/10.1016/j.system.2015.11.007>
- Zhou, S. (2018). “You just picked it up.” *The influence of informal language contact on the phrasal verb knowledge of international students in the UK* [Master dissertation]. University of Cambridge.
- Zhou, S., & Briggs Baffoe-Djan, J. G. (2022). “You just picked it up.” The informal language contact and phrasal verb knowledge among international students in the UK. *Study Abroad Research in Second Language Acquisition and International Education*.
- Zhou, Y., & Todman, J. (2009). Patterns of adaptation of Chinese postgraduate students in the United Kingdom. *Journal of Studies in International Education*, 13(4), 467–486. <https://doi.org/10.1177/1028315308317937>

Appendix A Ethics approval

CUREC confirmation



Joshua McGrane

Fri 07/06, 15:21



Dear Siyang,

Title: "Informal language contact and phrasal verb development among international students studying in the UK"

The above application (reference number ED-CIA-19-216) 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.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse

http://www.nspcc.org.uk/Inform/research/questions/reporting_child_abuse_wda74908.html

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Yours sincerely,

Joshua McGrane

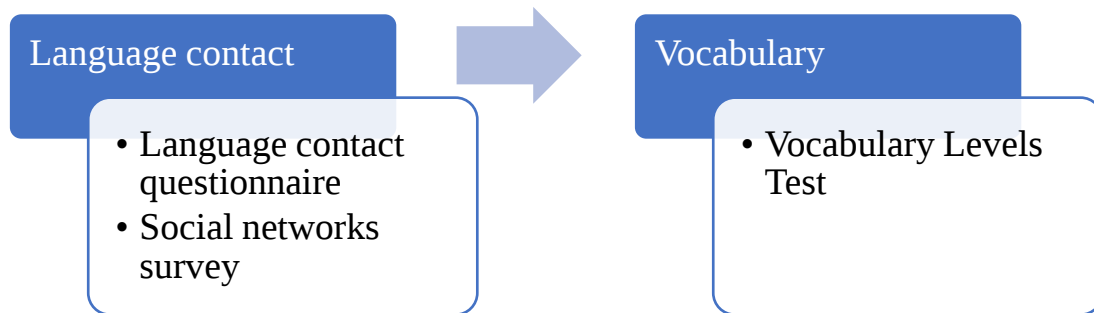
Member of DREC

...

Appendix B Research information sheet and consent form

Informal language contact and vocabulary development among international students studying in the UK

Siyang Zhou, Department of Education, University of Oxford



Background: Out-of-class language contact during study abroad has been found to have significant relationship with various linguistic competence, such as oral fluency and pragmatics. However, international students tend to have limited informal language contact. Research shows that vocabulary is often a strong predictor of academic achievement of international students. Thus, the present study aims to find out whether the social interaction of international students will have an impact on their general vocabulary and phrasal verb development over one academic year.

Significant of the study: 1. Big sample size and long tracking period in study abroad research. 2. Investigates the role of social networks on vocabulary development. 3. Provide recommendations to practitioners/study-abroad students.

Instruments samples:

	
Vocabulary Levels Test 3000 & 4000	Productive phrasal verb test
	
Language contact questionnaire	Social networks survey

INTERVIEW INFORMED CONSENT FORM

Dear participant:

My name is Siyang Zhou and I am a researcher from the Department of Education of the University of Oxford.

You are sincerely invited to participate in the research entitled “informal language contact and vocabulary development of international students in the UK”. This study involves a vocabulary test and two language contact questionnaires, which will be conducted three times a year.

Thank you for participating in the follow-up interview about your social life during study abroad. It is guaranteed that all the data will be confidential and your privacy will be strictly protected. Any information which might potentially identify you will NOT be used in published material. You can withdraw from the research anytime. Your support and participation will be much appreciated.

Siyang Zhou

Nov 14th 2019

By signing below, you are agreeing that: 1. I have read and understood the purposes of the study. 2. I understand I can withdraw at any time. 3. I agree to participate in the study as outlined to me.

Participant Name: _____ Date: _____

Appendix C Participant Recruitment Presentation

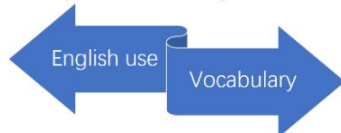
Study-abroad research by



&



• Recap of this study:



- Collect data three times in a year.
- Amazon voucher lucky draw at each time.
- Free VIP tours in Oxford!

• Remember me? Siyang Zhou

- PhD student at Oxford University
- IELTS Score: Overall 8.5



What's new with you?

- Have you used English more?
- Have you made more English-speaking friends?
- Have you learnt any new vocabulary?



SOCIAL NETWORK SURVEY TIPS:

Please give the information of up to 5 closest English-speaking friends you have.

- Write about yourself
- Write about friends from your home country
- Only write about British friends
- Write about friends not from your country but you speak English with.



STEP 1: ON PAPER

- Vocabulary Test PART 1 & PART 2 (15 min)

- Rules:
- No cell phones. No talking.
- Please try your best! 😊



STEP 2: ON MOBILE PHONE

- PART 3 3 min
- Language contact questionnaire
- <https://bit.ly/37Zw1A6>
- PART 4 7 min
- Social network survey
- <https://bit.ly/2yW5QKx>



STEP 3: ON PAPER

- Vocabulary Test PART 3 (10 min)

- No cell phones. No talking.



Appendix D Test papers

STUDY ABROAD RESEARCH - TIME 1 (40 minutes)

Student Number: CEG _____

PART 1 Read each sentence carefully, and write what you think the missing words (a verb and a preposition) are. (10 minutes)

E.g. I can't hear what you are saying. Can you t_____ that music o_____? (*stop by using a switch*)

Answer: *turn off*

1. C_____ o_____, don't be shy and tell us your story! (*said as encouragement*)
2. She opened the door and let him c_____ i_____. (*entered*)
3. Our car b_____ d_____ yesterday. (*stopped working*)
4. I know you're tired, but we can't stop now. We have to g_____ o_____ until we finish. (*continue*)
5. After dinner, he t_____ her b_____ to her house. (*returned*)
6. Unfortunately we've r_____ o_____ of biscuits. (*used completely*)
7. He t_____ u_____ to the meeting half an hour late. (*arrived, appeared*)
8. The fire has finally been p_____ o_____. (*extinguished*)
9. He r_____ o_____ for the glass on the table. (*stretched an arm so as to grab*)
10. We were h_____ u_____ by heavy traffic. (*delayed*)
11. Take the carrots and c_____ the ends o_____. (*remove*)
12. I don't want to stay in and cook tonight. Let's g_____ o_____ for a meal. (*leave your house for a special reason*)
13. She h_____ b_____ the laughter with great effort. (*contained, repressed*)
14. When we first met, we didn't like each other much. Now we g_____ o_____ really well. (*have a good relationship, be friends*)
15. I could barely m_____ o_____ his face in the dark. (*see*)
16. I am trying to get Peter to tell me when he wants to go on holiday, but it's very difficult to p_____ him d_____ to an exact date. (*make him decide*)
17. I think I t_____ a_____ my mother. We look very similar. (*similar to, be like*)
18. Excellent product quality is what made the brand s_____ o_____ from its competitors. (*distinguish itself by being better*)
19. Take the bus and g_____ o_____ the bus at the third stop. (*leave the bus*)
20. I didn't notice that the button c_____ o_____ my shirt. (*be detached, separate from*)

-----Well done! Please turn to the next page-----

PART 2 Please tick the word that goes with each meaning. (5 minutes)

Example:

	game	island	mouth	movie	song	yard
land with water all around it		✓				
part of your body used for eating and talking			✓			
piece of music					✓	

21.

	gown	maid	mustache	paradise	pastry	vinegar
hair on your upper lip						
perfect place						
small baked food						

22.

	asthma	chord	jockey	monk	rectangle	vase
container for cut flowers						
group of musical notes that are played at the same time						
shape with two long and two short sides						

23.

	batch	dentist	hum	lime	pork	scripture
green fruit						
low, constant sound						
meat from pigs						

24.

	amnesty	claw	earthquake	perfume	sanctuary	wizard
liquid that is made to smell nice						
man who has magical powers						
safe place						

25.

	altitude	diversion	hemisphere	pirate	robe	socket
height						
kind of clothing						
person who attacks ships						

26.

	applaud	erase	jog	intrude	notify	wrestle
announce						
enter without permission						
remove						

27.

	bribe	expire	immerse	meditate	persecute	shred
cut or tear into small pieces						
end						
think deeply						

28.

	commemorate	growl	ignite	pierce	renovate	swap
catch fire						
exchange						
go into or through something						

29.

	bald	eternal	imperative	lavish	moist	tranquil
calm and quiet						
having no hair						
slightly wet						

30.

	diesel	incidental	mandatory	prudent	superficial	tame
not dangerous						
required						
using good judgment						

Student number: CEG _____

Please take out your phone now.

PART 3 (7 minutes)

Online language contact questionnaire

<https://bit.ly/2NJyCp4>



PART 4 (7 minutes)

Online social network survey

<https://bit.ly/2yW5QKx>



Please put away your phone now.

PART 5 Please circle a verb and a preposition from the table to find the best phrasal verb for each sentence. (10 minutes)

E.g. *We tried to buy tickets for the concert but they _____ in a few hours.*

Finished	On
Run	Out
Sold	Up

31. Our car _____ yesterday.

Broke	Along
Brought	Down
Kept	Off

32. _____, don't be shy and tell us your story!

Calm	Down
Come	On
Get	Up

33. He _____ to the meeting half an hour late.

Made	Along
Took	In
Turned	Up

34. The fire has finally been _____.

Keep	Down
Put	Off
Stop	Out

35. Take the carrots and _____ the ends _____.

Cut	Down
Make	Off
Set	Up

36. I didn't notice that the button _____ my shirt.

Broke	Away
Came	Off
Took	Out

37. Take the bus and _____ the bus at the third stop.

Climb	Away
Get	Down
Go	Off

38. She opened the door and let him ____.

Come	Down
Start	In
Turn	Over

39. I know you're tired, but we can't stop. We have to ____ until we finish.

Come	On
Go	Off
Take	Up

40. After dinner, he ____ her ____ to her house.

Gave	Away
Handed	Back
Took	Up

41. Unfortunately we've ____ of biscuits.

Gone	Around
Run	Off
Taken	Out

42. She ____ the laughter with great effort.

Held	Back
Kept	Down
Put	Off

43. He ____ for the glass on the table.

Made	In
Put	Out
Reached	Over

44. Excellent product quality is what made the brand ____ from its competitors.

Show	Around
Stand	Out
Turn	Up

45. I think I ____ my mother. We look very similar.

Bring	After
Put	Along
Take	Up

46. We were ____ by heavy traffic.

Held	Away
Moved	Off
Turned	Up

47. I could barely ____ his face in the dark.

Look	In
Make	Off
Take	Out

48. When we first met, we didn't like each other much. Now we ____ really well.

Bring	Away
Get	On
Set	Over

49. I am trying to get Peter to tell me when he wants to go on holiday, but it's very difficult to ____ him ____ to an exact date.

Get	Down
Pin	Out
Put	Up

50. I don't want to stay in and cook tonight. Let's ____ for a meal.

Come	Back
Go	Over
Sit	Out

Thank you for your participation!

Appendix E the Language Contact Questionnaire

How often do you do the following activities in your spare time in English?

1. Read English newspaper, books, magazines, and academic materials
2. Read English Internet webpages on the computer or mobile phones
3. Read English timetables, menus, maps, announcements, etc.
4. Use social media in English (Facebook, Twitter, WhatsApp, WeChat, etc.)
5. Use social media **in your first language** (Facebook, Twitter, WhatsApp, WeChat, etc.)
6. Watch TV in English
7. Watch movies in English
8. Watch video websites in English (YouTube, Netflix)
9. Playing computer games in English
10. Listen to English songs
11. Listen to English radio/podcast
12. Listen to people's conversations in English
13. Write homework in English
14. Write emails in English
15. Talk with friends from your country **in your first language**
16. Talk with international friends in English
17. Talk with British friends in English
18. Talk to service personnel (e.g., waiters, shop assistants) in English

Appendix F IELTS speaking test questions

Time 1

Part 1

Studies

I'd just like to ask you some questions about your studies.

What are you studying?

Why did you choose that particular course?

What is your favourite subject?

Are you friends with many other students on your course?

What job would you like when you have completed all your studies?

Will you have to do further studies in the future?

Free time

How much free time do you have per week?

How do you like to relax at the end of the day?

What kinds of things do you enjoy doing at the weekends?

Do you have any plans for next weekend?

Do you prefer to spend your free time alone or with other people?

Do you have enough free time to do all the things you enjoy?

What would you do if you had more free time?

Part 2

Describe an interesting journey you have been on.

You should say:

where you went

how you travelled there

who you went with

and explain what was so memorable about the journey.

Part 3

Healthy lifestyle

What are the most popular ways of keeping healthy in your country?

Do you think most people worry more about their health as they get older?

Why do you think some people continue bad habits when they know that they are damaging to their health?

How can children be encouraged to adopt healthy eating habits?

Do you think people have become more health conscious in recent years?

Could governments do more to promote healthier lifestyle options?

Time 2

PART 1

Hometown

I'd just like to ask you some questions about your hometown.

Where is your hometown?

What was it like growing up there?

Has it changed much since you were a child?

What do you like most about living there?

What kinds of things can visitors to your hometown go and see?

Do you think you will live there when you are older?

Hobbies

Do you have a hobby?

What equipment do you need for it?

Do you think hobbies should be shared with other people?

Did you have a hobby as a child?

What hobbies are popular in your country?

Why do you think people have hobbies?

Part 2

Describe an interesting journey you have been on.

You should say:

where you went

how you travelled there

who you went with

and explain what was so memorable about the journey.

Part 3

Leisure activities

What types of leisure activities are popular in your country?

Why it is important for people to have time for leisure activities?

Why are some activities more popular than others?

Are the types of leisure activities that are popular today the same as those that were popular when your parents were young?

What types of leisure activities may become more popular in the future?

Time 3

Part 1

Evenings

What do you often do in the evenings?

Do you do the same thing every evening?

Do you prefer to spend your evenings with family or friends?

Do you ever work or study in the evenings?

What is a popular activity for young people in your country in the evenings?

Do you do the same thing in the evenings as you did when you were a child?

Family and friends

Do you spend much time with your family?

Who are you closest to in your family?

Do you prefer spending time with your family or friends?

Who is your best friend?

Are you still friends with people from your childhood?

Part 2

Describe a film you found interesting.

You should say:

when you saw this film

why you decided to see this film

what happened in the film

and explain why you found this film interesting.

Part 3

Travel

What methods of transport are most popular in your city / country?

What are some of the problems of long-distance travelling?

Has the way people travel changed much in the last few decades?

Do you agree that governments should do more to encourage greener methods of transport?

What measures could be taken to reduce problems of congestion in cities?

Do you think that transport problems are worse in urban or rural areas?

Appendix G Interview schedules

Time 1

Introduction

1. Why do you choose to study abroad?

Informal language contact

2. Tell me about your living situation here:
 - a. Where do you live?
 - b. Who do you live with?
 - c. Which languages are used in your home here?
3. Overall, do you use English more than your first language outside the classroom?
 - a. What do you usually use English for?
 - b. What do you usually use your first language for?
4. Please tell me more about an activity that you use English most often outside the classroom.
5. Do you feel that you are using English as much as you expected to?

Social network

6. Tell me about your friends here in the UK:
 - a. Do you have friends from your course, from other courses at university, or from outside the university?
 - b. Where are these friends from (i.e., international students, or local people)?
 - c. What language(s) is/are used with these friends?
 - d. What do you do with your English-speaking friends?
 - e. What do you usually talk about with your English-speaking friends?
7. Can you tell me about how you became friends with one of the persons you listed in the social network survey?

Vocabulary learning

8. Do you pay attention to new vocabulary when you are not in class? If yes, what kind of vocabulary do you tend to notice? Why?
9. Do you think about whether you used certain words/expressions right? Does that help you remember the words better?
10. What kind of activities help you learn vocabulary? Do you think having English-speaking friends helps you improve English vocabulary?
11. Have you learnt any new words/phrases recently? Could you give me some examples?
12. Are you happy about your language contact and social networks at the moment?

Time 2

Introduction

1. How is your study going?
2. Did you go back to your home country during the holiday? If yes, how long did you stay? If not, what did you do during class vacation?

Informal language contact

3. Did your living situation change?
4. Do you use English or your L1 more now? Some examples?

Social network

5. Do you make more friends in the past three months? (L1 or English-speaking?)
6. Do you make any British friends?
7. Do you think students from the same country tend to hang out together? Why or why not?
8. What are the benefits of having L1-speaking friends/English-speaking friends?
9. What are the difficulties of making English-speaking friends?
10. Do you prefer having native English speaker friend or non-native English speaker friend? Why
11. What's your ideal ratio of L1-speaking and English-speaking friends?
12. In your opinion, what kind of personality attracts English-speaking friends?

Vocabulary learning

13. Are you trying hard to learn more vocabulary/phrases at the moment?
14. Have you learnt any new words/phrases recently? Could you give me some examples?
15. Are you happy about your language contact and social networks at the moment?

Time 3**Introduction**

1. What's the biggest change after two months?

Language contact and social network

2. Do you use English or your L1 more?
3. What's your main source of English contact?
4. Which friend do you contact most often this month?
5. Do you keep in contact with your English-speaking friends? Why or why not?
6. Do you prefer to make international friends of the same gender or not?
7. Do you think COVID-19 has an impact on your English learning and friendship with foreign friends?

Phrasal verb learning

8. What's your opinion on the difficulty of phrasal verbs?
9. Do you think phrasal verbs are useful?
10. When do you notice phrasal verbs in your life?
11. How do you learn new phrasal verbs? From class or from daily life? Can you give me some examples?
12. Do you think studying abroad helps you to improve phrasal verbs? Why or why not?
If yes, how long does it take to see the gains?
13. Show the interviewee his PV test scores, LCQ scores, and SASNS scores and ask him to explain the possible reasons for the changes.

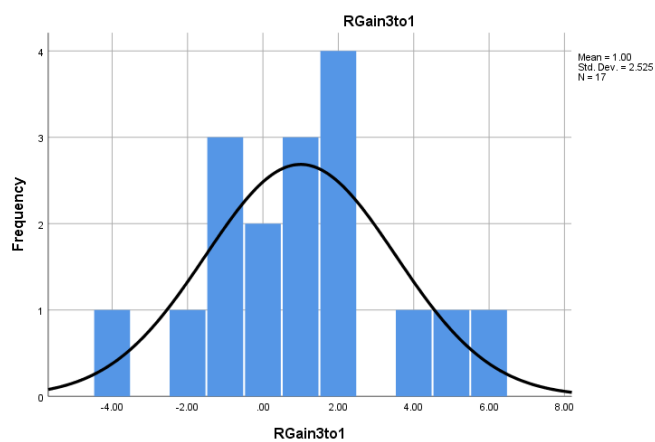
Appendix H Quantitative analysis for Finding Chapter IV

H1. Descriptive statistics of PV gains

Time 1 to 3 PV gains

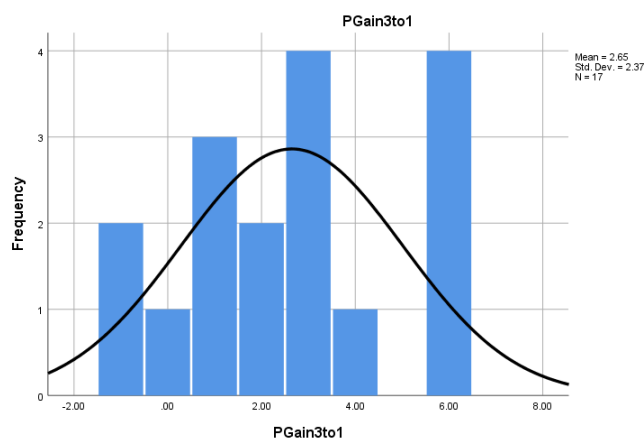
The mean gain between Time 1 and Time 3 receptive tests was 1.00 (N=17, SD=2.52), with a minimum value of -4 and a maximum value of 6. It reached normal distribution.

Histogram of Time 1 to Time 3 receptive PV test gains



The mean gain between Time 1 and Time 3 productive tests was 2.65 (N=17, SD=2.37), with a minimum value of -1 and a maximum value of 6. It did not reach normal distribution visually. Square root and log transformation were tried, but did not reach better histogram shape, so this variable was not transformed.

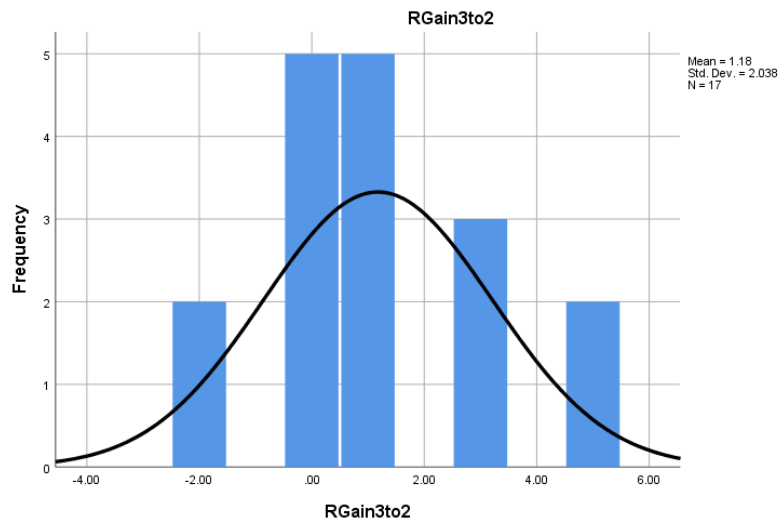
Histogram of Time 1 to Time 3 productive PV test gains



Time 2 to 3 PV gains

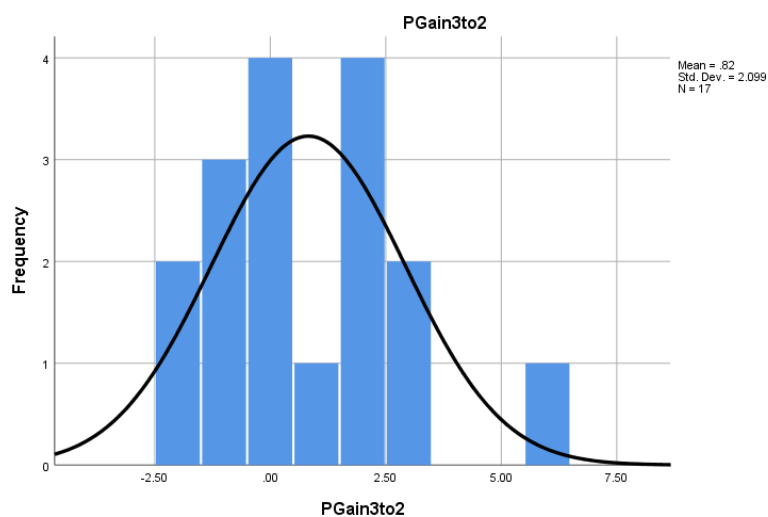
The mean gain between Time 2 and Time 3 receptive tests was 1.177 (N=17, SD=2.04), with a minimum value of -2 and a maximum value of 5. It reached normal distribution, despite some huge gaps in the histogram. I

Histogram of Time 2 to Time 3 receptive PV test gains



The mean gain between Time 2 and Time 3 productive tests was .824 (N=17, SD=2.09), with a minimum value of -2 and a maximum value of 6. Since there was no suitable technique to transform this histogram pattern, when doing the inferential stats, non-parametric tests were carried out alongside parametric tests.

Histogram of Time 2 to Time 3 productive PV test gains



H2. Independent-samples t-test on semantic opacity

Comparison of PV scores of literal meanings and figurative meanings

Test	Semantic transparency	N	Mean	SD	t	p
T1 receptive	literal	7	0.67	0.27	.808	.429
	figurative	13	0.5	0.27		
T2 receptive	literal	7	0.65	0.23	.493	.628
	figurative	13	0.49	0.27		
T3 receptive	literal	7	0.73	0.3	.782	.444
	figurative	13	0.6	0.35		
T1 productive	literal	7	0.51	0.29	1.137	.27
	figurative	13	0.35	0.31		
T2 productive	literal	7	0.49	0.29	.927	.366
	figurative	13	0.35	0.34		
T3 productive	literal	7	0.59	0.3	.851	.406
	figurative	13	0.46	0.35		
RGainT2-T1	literal	7	-0.02	0.19	.297	.770
	figurative	13	0.00	0.12		
RGainT3-T1	literal	7	0.01	0.14	.822	.422
	figurative	13	0.06	0.12		
RGainT3-T2	literal	7	0.03	0.16	.473	.642
	figurative	13	0.06	0.12		
PGainT2-T1	literal	7	0.02	0.12	1.047	.309
	figurative	13	0.08	0.11		
PGainT3-T1	literal	7	0.06	0.08	1.345	.195
	figurative	13	0.12	0.10		
PGainT3-T2	literal	7	0.04	0.06	.189	.852
	figurative	13	0.05	0.07		

Note: P=productive, R=receptive

H3. PV produced in the learner corpus

T1PV	Frequency	T2PV	Frequency	T3 PV	Frequency
go out	14	grow up	11	go out	15
focus on	11	go out	8	focus on	6
get up	8	go back	7	go back	5
stay up	6	hang out	6	depend on	4
depend on	3	depend on	4	grow up	4
go back	3	get up	4	slow down	3
grow up	3	wake up	3	come up	2

calm down	2	come back	2	get along	2
carry on	2	give up	2	sit down	2
cut down	2	go in	2	catch up	1
put in	2	go up	2	check in	1
base on	1	put in	2	come out	1
bright up	1	bump into	1	die out	1
build up	1	carry on	1	fight back	1
cheer up	1	catch up	1	find out	1
come back	1	come in	1	get into	1
come in	1	come out	1	get together	1
come out	1	find out	1	give up	1
find out	1	focus on	1	go down	1
get back	1	get together	1	go into	1
give up	1	go around	1	hurry up	1
hang out	1	go on	1	keep up	1
lie down	1	go over	1	move out	1
look after	1	make up	1	open up	1
slide down	1	pick up	1	pass away	1
speed up	1	stay up	1	put in	1
take up	1	walk around	1		
walk around	1	write down	1		

H4. Spearman correlation between total words and PV number

	T1PVtype			T1PVtoken		
	ρ	p	N	ρ	p	N
T1totalwords	0.268	0.267	19	0.177	0.47	19

	T2PVtype			T2PVtoken		
	ρ	p	N	ρ	p	N
T2totalwords	.614**	0.003	21	.676**	0.001	21

	T3PVtype			T3PVtoken		
	ρ	p	N	ρ	p	N
T3totalwords	.626*	0.017	14	.749**	0.002	14

H5. PV number change between two time points and three time points

Repeated-measures ANOVA

Variable	df	F	p	Partial η^2
PV types	2	1.024	.407	.226
PV tokens	2	2.209	.180	.387

Friedman's test

Variables	N	Chi-Square	df	Asymp. Sig.
PV types	9	2.69	2	0.261
PV tokens	9	3.161	2	0.206

H6. PV sophistication change between two time points and three time points

Wilcoxon Signed Ranks Test

	T1-T2	T2-T3	T1-T3
Z	-.416 ^b	.000 ^c	-1.483 ^d
Asymp. Sig. (2-tailed)	.677	1.000	.138
a. Wilcoxon Signed Ranks Test			
b. Based on negative ranks.			
c. The sum of negative ranks equals the sum of positive ranks.			
d. Based on positive ranks.			

Friedman's Test

N	6
Chi-Square	1.733
df	2
Asymp. Sig.	.420

H7. PV semantic accuracy change between two time points and three time points

Wilcoxon Signed Ranks Test

	T1-T2	T2-T3	T1-T3
Z	-.135 ^b	-.447 ^b	-1.000 ^c
Asymp. Sig. (2-tailed)	.892	.655	.317
a. Wilcoxon Signed Ranks Test			
b. Based on negative ranks.			
c. Based on positive ranks.			

Friedman's Test

N	6
Chi-Square	.286
df	2
Asymp. Sig.	.867

H8. PV grammatical accuracy change between two time points and three time points

Wilcoxon Signed Ranks Test

	T1-T2	T2-T3	T1-T3
Z	-.807b	-.070b	-1.378c
Asymp. Sig. (2-tailed)	.420	.944	.168
a. Wilcoxon Signed Ranks Test			
b. Based on negative ranks.			
c. Based on positive ranks.			

Friedman's Test

N	6
Chi-Square	.095
df	2
Asymp. Sig.	.953

H9. Spearman correlation between learner corpus PV characteristics and PV test scores at Time 1, Time 2 and Time 3

Time 1 (n=16)

Variables	PPV		RPV	
	<i>ρ</i>	<i>p</i>	<i>ρ</i>	<i>p</i>
Type	-.062	.820	-.045	.870
Token	.064	.815	-.010	.972
Sophistication	.358	.190	.133	.636
Semantic accuracy	.482	.069	.432	.108
Grammatical accuracy	.387	.154	.348	.204

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Time 2 (n=19)

Variables	PPV		RPV	
	<i>ρ</i>	<i>p</i>	<i>ρ</i>	<i>p</i>
Type	.364	.126	.378	.122
Token	.339	.155	.360	.143
Sophistication	.270	.312	.263	.344
Semantic accuracy	-.231	.389	-.280	.312
Grammatical accuracy	.148	.585	.114	.685

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Time 3 (n=10)

Variables	PPV		RPV	
	ρ	p	ρ	p
Type	.524	.120	.526	.118
Token	.392	.263	.261	.466
Sophistication	.434	.244	.413	.270
Semantic accuracy	a	a	a	a
Grammatical accuracy	.764*	.016	.512	.159

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

H10. Spearman correlation between learner corpus PV characteristics and overall English proficiency at Time 1, Time 2 and Time 3

Time 1

Variables	VLT1 (n=15)		IELTS (n=19)	
	ρ	p	ρ	p
Type	-.018	.951	.037	.882
Token	-.145	.606	.083	.735
Sophistication	.101	.721	.096	.706
Semantic accuracy	.374	.170	.490*	.039
Grammatical accuracy	.096	.734	.040	.874

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Time 2

Variables	VLT2 (n=19)		IELTS (n=20)	
	ρ	p	ρ	p
Type	-.195	.424	.120	.614
Token	-.120	.625	.190	.423
Sophistication	.464	.070	.519*	.033
Semantic accuracy	-.197	.464	-.352	.165
Grammatical accuracy	-.046	.866	-.126	.629

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Time 3

Variables	IELTS (n=13)	
	ρ	p
Type	.235	.439
Token	.210	.492
Sophistication	.091	.777

Semantic accuracy	.455	.138
Grammatical accuracy	.046	.887

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

H11. Speaking topics and learner corpus PV characteristics

Kruskal-Wallis Test

	Sophistication	Semantic	Grammar	Type
Kruskal-Wallis H	4.044	1.080	1.172	.256
df	2	2	2	2
Asymp. Sig.	.132	.583	.556	.880

H12. Task types and learner corpus PV characteristics

Mann-Whitney U Tests results of different task types

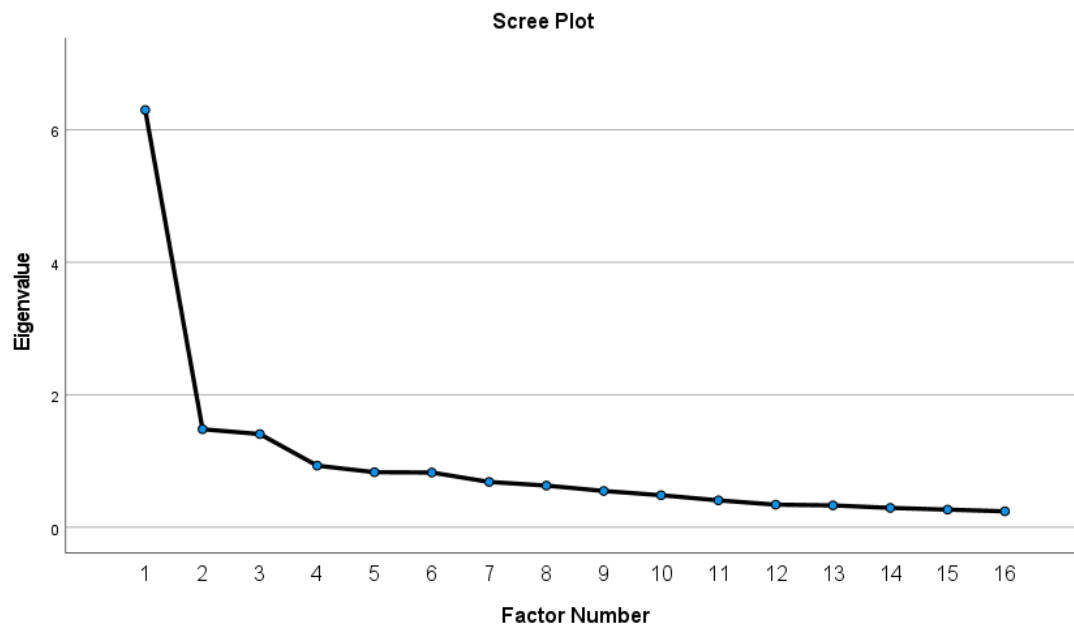
Variable	Sophistication n	PV type	Semantic accuracy	Grammatical accuracy
Total N	9	9	9	9
Mann-Whitney U	18.000	1.000	10.500	7.500
Wilcoxon W	24.000	7.000	16.500	13.500
Test Statistic	18.000	1.000	10.500	7.500
Standard Error	3.857	3.873	3.841	3.857
Standardized Test Statistic	2.334	-2.066	.391	-.389
Asymptotic Sig.(2-sided test)	.020	.039	.696	.697
Exact Sig.(2-sided test)	.024	.048	.714	.714

Appendix I Quantitative analysis for Finding Chapter V

I1. T1-T2 significant changes of LCQ items before Bonferroni correction

Item		Mean	SD	t	df	N	p
Use social media in English	T1	3.03	1.527	2.20	99	100	.030
	T2	2.69	1.631				
Speaking English with service personnel	T1	3.75	1.350	2.73	98	99	.007
	T2	3.35	1.350				

I2. Time 2 EFA Scree Plot



I3. The significant relationships between L1 and LCQ scores

Time 1 independent-samples t-test results

Group Statistics					
Chinese L1		N	Mean	SD	SE Mean
Books	1	123	1.72	0.99	0.09
	0	48	2.31	1.32	0.19
Websites	1	123	2.20	1.00	0.09
	0	48	3.23	1.42	0.21
Timetables	1	123	2.75	1.03	0.09

	0	48	3.60	1.16	0.17
Social media L2	1	123	2.82	1.42	0.13
	0	48	3.88	1.42	0.21
SocialmediaL1	1	123	3.28	1.34	0.12
	0	48	1.94	1.68	0.24
TV	1	123	1.84	1.18	0.11
	0	48	3.00	1.40	0.20
Movie	1	123	2.67	1.07	0.10
	0	48	3.54	1.22	0.18
YouTube	1	123	2.40	1.26	0.11
	0	48	3.48	1.27	0.18
Game	1	71	2.00	1.46	0.17
	0	23	3.26	1.63	0.34
Songs	1	122	3.29	1.22	0.11
	0	48	3.88	1.18	0.17
Conversation	1	122	2.70	1.06	0.10
	0	48	3.58	1.01	0.15
Emails	1	122	2.81	1.23	0.11
	0	48	3.40	1.20	0.17
International friends	1	122	3.28	1.38	0.13
	0	48	4.15	1.03	0.15

1=Chinese, 0=Non-Chinese

	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval	Lower	Upper
Books*	-2.795	68.332	.007	-.589	.211	-1.009	-.169	
Websites	-4.588	66.048	.000	-1.026	.224	-1.472	-.579	
Timetables	-4.712	169	.000	-.856	.182	-1.215	-.497	
Social media	-4.357	169	.000	-1.054	.242	-1.531	-.576	
Social media L1	5.486	169	.000	1.347	.246	.862	1.832	
TV	-5.500	169	.000	-1.163	.211	-1.580	-.745	
Movie	-4.621	169	.000	-.875	.189	-1.249	-.501	
YouTube	-5.029	169	.000	-1.081	.215	-1.505	-.657	
Game	-3.491	92	.001	-1.261	.361	-1.978	-.544	
Songs*	-2.851	168	.005	-.588	.206	-.995	-.181	
Conversation	-4.980	168	.000	-.887	.178	-1.238	-.535	
Emails*	-2.811	168	.006	-.584	.208	-.995	-.174	
International friends	-4.462	114.512	.000	-.867	.194	-1.252	-.482	

*Non-significant after Bonferroni correction

The rest are significant after Bonferroni correction

Time 2

Chinese L1		N	Mean	SD	SE Mean
Timetable	1	118	2.86	1.303	0.120
	0	32	3.66	1.066	0.188
Social media L2	1	118	2.51	1.518	0.140
	0	32	3.91	1.445	0.255
L1 social media	1	118	3.86	1.207	0.111
	0	32	2.28	1.708	0.302
TV	1	118	1.93	1.218	0.112
	0	32	3.41	1.241	0.219
Movies	1	118	2.53	1.231	0.113
	0	32	3.47	1.414	0.250
YouTube	1	118	2.41	1.302	0.120
	0	32	3.28	1.250	0.221
Song	1	118	3.07	1.407	0.129
	0	32	3.91	1.174	0.208
Radio	1	118	1.15	1.067	0.098
	0	32	2.19	1.533	0.271
Email	1	118	2.76	1.145	0.105
	0	32	3.63	0.907	0.160
International friends	1	118	2.64	1.285	0.118
	0	32	4.00	1.503	0.266
British friends	1	118	2.23	1.476	0.136
	0	32	3.31	1.839	0.325
Service personnel	1	118	3.25	1.247	0.115
	0	32	4.03	1.448	0.256

1=Chinese, 0=Non-Chinese

	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval	
						Lower	Upper
Timetable	-3.195	148	.002	-.800	.250	-1.295	-0.305
Social Media	-4.667	148	.000	-1.398	.300	-1.990	-0.806
L1social Media	4.893	39.774	.000	1.575	.322	0.924	2.225
TV	-6.050	148	.000	-1.474	.244	-1.956	-0.993
Movies	-3.688	148	.000	-.935	.253	-1.436	-0.434
YouTube	-3.397	148	.001	-.874	.257	-1.383	-0.366
Song	-3.091	148	.002	-.838	.271	-1.375	-0.302
Radio	-3.590	39.500	.001	-1.035	.288	-1.618	-0.452
Email	-3.936	148	.000	-.862	.219	-1.295	-0.429
International friends	-5.103	148	.000	-1.356	.266	-1.881	-0.831
British friends	-3.075	42.427	.004	-1.084	.352	-1.795	-0.373
Service personnel	-3.052	148	.003	-.785	.257	-1.294	-0.277
Timetable	-3.195	148	.002	-.800	.250	-1.295	-0.305

*Non-significant after Bonferroni correction

The rest are significant after Bonferroni correction

14. The relationship between overall English proficiency and LCQ scores

Time 1 Pearson correlation between IELTS, VLT and LCQ items

LCQ items	IELTS (n=128)		VLT1 (n=134)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Books	.289**	0.001	.254**	0.003
Websites	.290**	0.001	.394**	0
Timetables	0.167	0.060	.226**	0.009
English social media	0.037	0.680	0.137	0.114
L1 social media	-0.028	0.751	-.253**	0.003
TV	.181*	0.041	.290**	0.001
Movie	.206*	0.020	.303**	0
YouTube	0.102	0.252	.315**	0
Game	0.144	0.250	.385**	0.002
Songs	0.135	0.132	.236**	0.006
Conversation	.302**	0.001	.384**	0
Emails	.180*	0.042	.241**	0.005
Homework	0.038	0.673	0.053	0.541
L1 friends	0.112	0.208	-0.057	0.511
International friends	0.028	0.755	.203*	0.019
British friends	0.065	0.466	0.151	0.082
Service personnel	0.093	0.301	0.106	0.222

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Time 2 Pearson correlation between IELTS, VLT and LCQ items

LCQ items	IELTS (n=75)		VLT2 (n=142)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Books	.258*	0.025	.267**	0.001
Websites	.351**	0.002	.320**	0.000
Timetables	.310**	0.007	.273**	0.001
English social media	0.181	0.121	.176*	0.037
L1 social media	-0.101	0.389	-.200*	0.017
TV	.347**	0.002	.317**	0.000
Movie	.296*	0.010	0.125	0.138
YouTube	0.162	0.164	.223**	0.008
Game	0.118	0.312	0.018	0.829
Songs	0.130	0.268	0.072	0.392

Conversation	0.154	0.186	0.072	0.392
Radio	0.147	0.207	.244**	0.003
Emails	0.164	0.160	.324**	0.000
Homework	0.127	0.278	.178*	0.034
L1 friends	-0.093	0.426	0.008	0.921
International friends	0.179	0.125	0.157	0.061
British friends	0.020	0.863	0.046	0.588
Service personnel	0.122	0.295	0.058	0.494

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

I5. The relationship between age and LCQ scores

Age and total LCQ scores

	LCQ1			LCQ2		
	r	p	N	r	p	N
Age	.015	.841	171	.061	.552	97

Age and LCQ factors

Variables	T1				T2	
	Receptive factor	Speaking factor	Writing factor	Productive factor	Multimedia factor	Reading factor
r	.035	-.041	-.070	.019	-.022	.191
p	.647	.589	.362	.849	.828	.056
N	172	172	172	100	100	100

I6. Time 1 and Time 2 difference of SASNS variables

Wilcoxon Signed Rank test of T1 and T2 SASNS data

Variables	Z	Asymp. Sig. (2-tailed)
L1 network size	-.280 ^b	.779
English network size	-.129 ^c	.898
Network dispersion	-1.717 ^c	.086
Topic multiplexity	-1.141 ^c	.254
Activity multiplexity	-.022	.982
Network density	-1.338 ^c	.181
Network intensity	-2.286 ^c	.022*
Friends' English proficiency	-2.332 ^c	.020*
Face-to-face contact frequency	-2.355 ^b	.019*

Online contact frequency	-.229 ^c	.819
a. Wilcoxon Signed Ranks Test		
b. Based on negative ranks.		
c. Based on positive ranks.		

*Note: The intensity of friendship is lower at T2. The relatively L2 proficiency of friends is lower. The face-to-face contact frequency of the listed friends is higher.

17. The relationship between the accommodation and SASNS variables

Time 1

	Kruskal-Wallis H	df	Asymp. Sig.
L1 network size	3.716	2	0.156
English network size	3.656	2	0.161
Network dispersion	2.822	2	0.244
Topic multiplexity	7.829	2	0.020*
Activity multiplexity	0.579	2	0.749
Network density	10.532	2	0.005
Network intensity	0.821	2	0.663
Friends' English proficiency	3.612	2	0.164
Face-to-face contact frequency	1.030	2	0.597
Online contact frequency	0.145	2	0.930

*Those living with people from different countries had higher topic multiplexity in L2.

Time 2

	Kruskal-Wallis H	df	Asymp. Sig.
L1 network size	3.233	2	0.199
English network size	3.321	2	0.190
Network dispersion	4.329	2	0.115
Topic multiplexity	2.959	2	0.228
Activity multiplexity	0.577	2	0.749
Network density	.548	2	0.760
Network intensity	2.986	2	0.225
Friends' English proficiency	3.276	2	0.194
Face-to-face contact frequency	1.677	2	0.432
Online contact frequency	6.680	2	0.035*

*Those living with people from different countries had higher online contact frequency with L2 friends.

18. The relationship between overall English proficiency and SASNS

Spearman correlation

IELTS	Time 1			Time 2		
	<i>ρ</i>	<i>p</i>	N	<i>ρ</i>	<i>p</i>	N
L1 network size	0.049	0.664	81	0.054	0.687	58
English network size	0.056	0.629	77	0.206	0.136	54
Network dispersion	-0.051	0.700	59	0.024	0.873	46
Topic multiplexity	.237*	0.032	82	0.151	0.259	58
Activity multiplexity	0.008	0.946	69	0.072	0.548	71
Network density	.339**	0.004	71	.362*	0.008	53
Network intensity	-.240*	0.044	71	-0.072	0.610	52
Friends' L2 proficiency	0.101	0.406	70	-0.122	0.390	52
Face-to-face frequency	-0.224	0.061	71	-0.203	0.148	52
Online contact frequency	-0.181	0.132	71	-0.042	0.770	52

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Spearman correlation

VLT	Time 1			Time 2		
	<i>ρ</i>	<i>p</i>	N	<i>ρ</i>	<i>p</i>	N
L1 network size	-0.021	0.815	123	-.317**	0.000	132
English network size	.187*	0.047	113	-0.114	0.211	122
Network dispersion	0.159	0.133	91	0.087	0.375	106
Topic multiplexity	0.143	0.107	128	0.172	0.050	130
Activity multiplexity	0.003	0.977	130	-0.047	0.659	90
Network density	0.088	0.379	103	.323**	0.000	114
Network intensity	0.086	0.394	100	0.000	0.998	112
Friends' L2 proficiency	-.284**	0.004	99	-0.011	0.910	112
Face-to-face frequency	-0.021	0.832	100	-.244**	0.009	112
Online contact frequency	0.084	0.407	100	-0.037	0.702	112

19. The relationship between motivation and SASNS variables

Spearman correlation

Motivation	Time 1			Time 2		
	<i>ρ</i>	<i>p</i>	N	<i>ρ</i>	<i>p</i>	N
L1 network size	0.064	0.499	115	0.015	0.889	92
English network size	.214*	0.028	105	0.101	0.354	86
Network dispersion	.353**	0.001	83	-0.047	0.695	71
Topic multiplexity	0.076	0.413	118	-0.004	0.971	92
Activity multiplexity	-0.035	0.748	88	-.292**	0.003	99
Network density	-.002	0.998	93	-0.044	0.694	81
Network intensity	-0.028	0.791	93	0.045	0.691	80
Friends' L2 proficiency	-0.080	0.446	92	-0.082	0.472	80
Face-to-face frequency	.315**	0.002	93	.292**	0.009	80
Online contact frequency	.212*	0.041	93	0.199	0.076	80

I10. The relationship between age and SASNS variables

Spearman correlation

Age	Time 1			Time 2		
	ρ	p	N	ρ	p	N
L1 network size	-0.036	0.707	114	-0.189	0.075	90
English network size	-.393**	0.000	104	-.290**	0.008	83
Network dispersion	-.391**	0.000	83	-0.208	0.084	70
Activity multiplexity	-0.044	0.682	91	0.018	0.862	101
Topic multiplexity	0.027	0.772	117	0.058	0.588	90
Network density	.032	0.763	94	0.001	0.992	79
Network intensity	-0.134	0.198	94	-0.105	0.362	78
Friends' L2 proficiency	.240*	0.020	93	0.076	0.508	78
Face-to-face frequency	-0.122	0.242	94	0.011	0.924	78
Online contact frequency	-0.087	0.407	94	-0.028	0.807	78

I11. The relationship between living arrangements and SASNS variables

Time 1

	Kruskal-Wallis H	df	Asymp. Sig.
L1 network size	3.716	2	0.156
English network size	3.656	2	0.161
Network dispersion	2.822	2	0.244
Network density	10.532	2	0.005*
Network intensity	0.821	2	0.663
Friends' L2 proficiency	3.612	2	0.164
Face-to-face frequency	1.030	2	0.597
Online contact frequency	0.145	2	0.930
Activity multiplexity	0.579	2	0.749
Topic multiplexity	7.829	2	0.020*

Time 2

	Kruskal-Wallis H	df	Asymp. Sig.
L1 network size	3.233	2	0.199
English network size	3.321	2	0.190
Network dispersion	4.329	2	0.115
Network density	0.548	2	0.760
Network intensity	2.986	2	0.225
Friends' L2 proficiency	3.276	2	0.194
Face-to-face frequency	1.677	2	0.432

Online contact frequency	6.680	2	0.035*
Activity multiplexity	0.577	2	0.749
Topic multiplexity	2.959	2	0.228

I12. The relationship between center locations and SASNS variables

Time 1

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
L1 network size	1708.500	4409.500	-0.782	0.434
English network size	1564.000	3842.000	-0.061	0.951
Network dispersion	1002.500	1743.500	-0.192	0.848
Network density	989.500	1619.500	-1.773	0.076
Network intensity	801.500	2946.500	-2.550	0.011*
Friends' L2 proficiency	1057.000	3137.000	-0.479	0.632
Face-to-face frequency	820.000	2965.000	-2.453	0.014*
Online contact frequency	934.500	3079.500	-1.538	0.124
Activity multiplexity	1963.500	4108.500	-0.840	0.401
Topic multiplexity	2010.500	4936.500	-0.017	0.987

Time 2

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
L1 network size	2402.000	5328.000	-0.739	0.460
English network size	2177.500	4068.500	-0.084	0.933
Network dispersion	1320.000	2646.000	-1.691	0.091
Network density	1915.500	3568.500	-.137	0.891
Network intensity	1629.000	3225.000	-1.311	0.190
Friends' L2 proficiency	1840.000	4118.000	-0.195	0.846
Face-to-face frequency	1406.500	3684.500	-2.501	0.012*
Online contact frequency	1690.500	3286.500	-0.965	0.334
Activity multiplexity	2291.500	5066.500	-0.330	0.741
Topic multiplexity	2005.000	4283.000	-2.088	0.037*

I13. LCQ item change between three time points

One-way repeated-measures ANOVA

N=26	T1		T2		T3		ANOVA			
Variable	M	SD	M	SD	M	SD	df	F	p	Partial η^2
Speaking with British friends	3.12	1.796	3.00	1.625	2.12	1.751	2,24	5.090	.014	.298
Listen to people's conversation	2.69	1.123	3.12	1.107	2.38	1.267	2,24	4.271	.026	.262
Watching movies	2.62	1.267	2.65	1.236	3.15	1.255	2,24	5.282	.013	.306

I14. SASNS variable changes between Time 1, Time 2 and Time 3

Friedman's test

Variables	N	Chi-Square	df	Asymp. Sig.
L1 network size	12	1.937	2	0.380
English network size	10	0.667	2	0.717
Network dispersion	10	2.214	2	0.331
Network density	8	2.462	2	0.292
Network intensity	8	2.545	2	0.280
Friends' L2 proficiency	8	2.100	2	0.350
Face-to-face frequency	8	7.000	2	0.030*
Online contact frequency	8	2.111	2	0.348
Activity multiplexity	13	5.489	2	0.064
Topic multiplexity	11	1.897	2	0.387

*Significant before Bonferroni correction

Detail of significance

Ranks	
	Mean Rank
face2faceT1	2.63
face2faceT2	1.88
face2faceT3	1.50

Participants had significant lower face-to-face contact frequency with L2 friends at T3.

Appendix J Sample coding scheme of thematic analysis

Language contact

Subcategories	Pattern Codes	Open coding	Examples
Overview	L1/L2 ratio	L2>L1	<i>Maybe 60% English and 40% Chinese.</i>
	Expectation	Lower than expectation	<i>A little bit lower than my expectation. Because after class I don't have time to go to shops. I don't have much opportunity to use English.</i>
Attitudes	Satisfaction	Satisfied	<i>I am satisfied. Not as frequent as I imagine but I made a lot of foreign friends.</i>
	Importance of L2 contact	High importance	<i>It's important. They teach you something more native, like British slangs.</i>
Language use by context	At dorm	Low L2 use	<i>But I don't socialize with them a lot because everyone has different routine and we are not from the same university.</i>
	In public places	With foreigners	<i>For social activities it depends on whether there is any non-Chinese speaker on the day. If all of us speak Chinese, we speak Chinese. As long as there is one foreigner, we would use English.</i>
L1 use	Communication	Call parents	<i>Maybe for two hours a day. Because I chat with my parents. I call them daily.</i>
Receptive L2 contact	Academic	Reading books	<i>I think reading an English book is the most helpful. Because I see many words I don't know. I will check it in the dictionary the write down.</i>
	Leisure	Movies	<i>When I watch English movies, I sometimes read aloud after it.</i>
	Leisure	YouTube	<i>I love YouTube. When I go on YouTube, it's only English.</i>
	Academic	IELTS	<i>Because I am preparing for IELTS. I read IELTS, listen to BBC sometimes.</i>
	Academic	Emails	<i>Recently I wrote many emails to contact universities in English. But except for this I don't use English for emails</i>
Interactive L2 contact	Miscellaneous purposes	Shopping, asking directions	<i>When I go to the daily issue, like go to the market and sometimes I can't find the way, I will use English</i>
	Academic	Group discussion	<i>But we have group work, there is a foreigner in my group. When we discuss group work we use English.</i>

Social networks

Subcategories	Pattern Codes	Open coding	Examples
Overview	L1/L2 friends' ratio	L1>L2	<i>Chinese friends 7: international friends 3 is the current situation. Ideally, I want half half.</i>
	NS/NNS preference	NS	<i>To be honest native speakers. Because they can help me with my speaking.</i>
L1 friends	Reason	Similar ideas	<i>They have shared ideas and they feel more comfortable staying with each other...</i>
International friends	On campus	Classmates	<i>All my classmates. Because we have very full timetable and when I finish class I just go home and no other time to chat with other people and to meet other friends, so it's all my classmates.</i>
	At dorm	Flatmates	<i>I am not interested in other people. I even haven't seen other friends in my floor.</i>
	On campus	Teachers	<i>I speak to them in class actively. After class we discuss some current issues, or domestic issues. I enjoy time with them.</i>
	In public	Random strangers	<i>I have one, but we only talked one time in our canteen. He is a student in Birkbeck. He studies master of Law, and he studied Chinese. We just talked to each other but we never met each other again.</i>
	Virtual setting	Online friends	<i>I don't have some foreign friends in reality. But I have some in social media. Some in Germany, some in UK, some in USA. I know them after I come to London, but I just met them online. The app is called Hinge.</i>
British friends	Romance	Partner	<i>Last year I know a guy called Kieve. We were like friends before Christmas. After Christmas we become couple.</i>
	Others	Lost contact	<i>I have a local friend. Because my phone was stolen, so I lost her contact.</i>
Topics in English	Academic	Homework	<i>We say hi and ask if you have handed in the homework, the scores and any revision.</i>
	Home country	Home country culture	<i>he would ask me what's special in China and I would ask him about Cyprus</i>
	Hobbies and interests	Sports	<i>Because we both like sports. We talk about football games yesterday, our</i>

			<i>hobby and interests</i>
	Others	Beauty	<i>The hair color, the shampoo, and how to prevent hair falling</i>
Activities in English	Group work	Review	<i>We did the review together. So just for study, not for shopping.</i>
	Pubs/clubs	Pub, club	<i>Half to the pub and half to the club. I don't really like clubbing so I sometimes say no.</i>
	Meals	Dinner	<i>Just eating. We have eaten together almost every day in the first month.</i>
Difficulties	Racism	Microaggression	<i>Firstly, maybe due to recent political issue, Koreans they do not want to talk to Chinese first. You can feel their different attitude toward Chinese and other nationalities.</i>
	Personalities attracting L2 friends	Confident, enthusiastic	<i>Confident, enthusiastic and don't be too traditional.</i>
SASNS completion	Example of L2 friends	Friendship formed early	<i>Actually when the first day I come to London, when the taxi driver drop me off at the dorm. She is outside as well, she was dropped off by another taxi driver. We said hi to each other.</i>

Developmental profile	Change	Increase	<i>This term my English speaking improved and my reading skill enhanced a lot. I read so many articles, related documents.</i>
		Decrease	<i>Less English class this term. Only 3 sessions per week. We don't pay much attention to English course.</i>
		Stagnation	<i>I think it's my writing skill. Because we didn't have many tasks with writing and my writing score still keep the same.</i>

COVID impact

Subcategories	Pattern Codes	Open coding	Examples
Change	Difficulties and challenges	Isolation	<i>The biggest change is being totally isolated. I think all people are facing the same thing.</i>
	Positive side	Self-learning	<i>Maybe this kind of isolation created a chance for self-learning, for more time</i>

			<i>for self-learning.</i>
L2 use	Leisure	Housemate (partner)	<i>Because we can't go outside, only two people stay at home. we have each other to speak.</i>
	Academic	Online classes, homework	<i>Listen to the teacher or answer teacher's questions. Reading literature when I write essay.</i>
	Leisure	Movies, roommate	<i>I think movies and speaking with my roommate are the main part. I think the frequency is a little bit more.</i>
	Leisure	News	<i>I also read COVID-19 news on BBC website, to know what's happening</i>
	Leisure	YouTube/podcast	<i>I occasionally listen to English radio. Podcast on Spotify. Hosted of a British actor I like. The topic is art. "Talk Art".</i>
L1 use	Families and friends	Boyfriend	<i>My boyfriend in China. My bestie too. When I am bored we talk.</i>
Social networks	Most frequently-contacted person	L1 friends in the UK	<i>One is Candice and another classmate in my class, Susie. We have been in good relationship before the COVID-19.</i>
	Relationship with English-speaking friends	Decrease	<i>She went back slightly after me. But we did not talk recently. After she fell sick we didn't contact again.</i>
	Relationship with L1-speaking friends	Same	<i>Chinese friend we tried to share the pressures together and tell jokes inside the group chat. Actually, when we isolated, we don't contact each other more frequently than before.</i>

Note: This table is a non-exhaustive example of the codes, to show the categories of codes.

Appendix K Quantitative analysis for Finding Chapter VI

K1. Correlation between LCQ and PV scores

Pearson correlation between T1 LCQ items and PV scores

	Productive (n=134)		Receptive (n=128)	
	r	p	r	p
Books	.203*	.019	.269**	.002
Websites	.394**	.000	.332**	.000
Timetables	.293**	.001	.222*	.012
English social media	.083	.341	.164	.065
L1 social media	.044	.617	-.034	.706
TV	.142	.101	.208*	.018
Movie	.227**	.008	.242**	.006
YouTube	.231**	.007	.176*	.047
Game	.339**	.002	.230*	.049
Songs	.123	.158	.128	.150
Conversation	.255**	.003	.242**	.006
Emails	.198*	.022	.099	.266
Homework	.116	.185	.057	.525
L1 friends	.167	.054	.120	.176
International friends	.056	.522	-.005	.957
British friends	.035	.690	.018	.842
Service personnel	.127	.145	.138	.120

Note: see full variable names in Finding Chapter II

Pearson correlation between T1 language contact factors and PV test scores

	Productive (n=130)		Receptive (n=124)	
	r	p	r	p
Receptive contact	.236**	.007	.210*	.019
Speaking contact	.098	.265	.101	.266
Writing contact	.114	.196	.069	.445

Pearson correlation between T2 language contact and PV T1-T2 gains

	Productive gain (n=84)		Receptive gain (n=75)	
	r	p	r	p
Books	.196	.073	.066	.576
Websites	.252*	.021	.250*	.031
Timetables	.195	.076	.319**	.005
English social media	.216*	.048	.299**	.009
L1 social media	-.173	.115	-.208	.074

TV	.156	.156	.234*	.043
Movie	.058	.599	.245*	.034
YouTube	.185	.091	.215	.064
Game	-.157	.154	.212	.068
Songs	.001	.992	-.006	.957
Conversation	.032	.775	.032	.788
Radio	.324**	.003	.003	.977
Emails	.038	.732	.037	.754
Homework	.030	.788	.115	.324
L1 friends	-.208	.058	.052	.658
International friends	.065	.557	.144	.218
British friends	.121	.271	-.041	.729
Service personnel	-.049	.660	.125	.286

Pearson correlation between LCQ total scores and PV T1-T2 gains

	Productive gain			Receptive gain		
	r	p	N	r	p	N
LCQ1	.230*	.043	78	.042	.731	71
LCQ2	.156	.158	84	.230*	.047	75

K2. Correlation between social networks and PV scores

Spearman correlation between T1 SASNS and PV scores

	Productive			Receptive		
	ρ	<i>p</i>	N	ρ	<i>p</i>	N
L1 network size	.110	.280	98	.073	.485	94
English network size	-.018	.869	90	-.019	.859	86
Network dispersion	-.037	.756	71	-.032	.795	68
Topic multiplexity	.235*	.019	100	.183	.074	96
Activity multiplexity	.195	.083	80	.304**	.008	75
Network density	.124	.289	75	.116	.335	71
Network intensity	-.115	.325	75	-.243*	.041	71
Friends' L2 proficiency	.004	.974	75	.015	.902	71
Face-to-face frequency	-.010	.930	75	-.155	.198	71
Online contact frequency	.312**	.006	75	.010	.936	71

Spearman correlation between T2 SASNS and PV scores

	Productive			Receptive		
	ρ	<i>p</i>	N	ρ	<i>p</i>	N
L1 network size	-.211*	.015	134	-.166	.063	127
English network size	-.035	.696	125	.012	.899	118

Network dispersion	.024	.804	106	-.079	.435	99
Topic multiplexity	.184*	.035	132	.220*	.013	125
Activity multiplexity	-.023	.827	94	-.111	.304	88
Network density	.232*	.012	117	.206*	.031	110
Network intensity	-.061	.518	115	-.008	.938	108
Friends' L2 proficiency	.053	.576	115	.060	.540	108
Face-to-face frequency	-.225*	.016	115	-.261**	.006	108
Online contact frequency	-.012	.900	115	-.004	.968	108

Spearman correlation between T2 SASNS and T1-T2 PV gains

	Productive gain			Receptive gain		
	ρ	p	N	ρ	p	N
L1 network size	-.265*	.019	78	-.098	.422	70
English network size	-.053	.643	78	-.007	.953	70
Network dispersion	.065	.583	74	-.014	.907	67
Topic multiplexity	-.123	.284	78	-.221	.066	70
Activity multiplexity	.178	.120	78	-.023	.853	70
Network density	.010	.941	71	-.091	.539	64
Network intensity	-.026	.832	71	-.026	.840	64
Friends' L2 proficiency	.023	.850	71	.056	.660	64
Face-to-face frequency	-.028	.816	71	-.058	.647	64
Online contact frequency	.039	.749	71	.174	.170	64

K3. Correlation between PV learner corpus and language contact

Time 1

(N=17)	Sophistication		Number (type)		Grammatical accuracy		Semantic accuracy	
	ρ	p	ρ	p	ρ	p	ρ	p
LCQ1	-.144	.582	-.176	.486	-.014	.956	.059	.821
Receptive contact	-.100	.704	-.284	.253	.024	.927	.125	.632
Speaking contact	-.145	.579	.030	.905	-.004	.989	-.217	.402
Writing contact	-.121	.644	.081	.750	-.378	.135	.066	.802

Time 2

(N=17)	Sophistication		Number (type)		Grammatical accuracy		Semantic accuracy	
	ρ	p	ρ	p	ρ	p	ρ	p
LCQ2	.146	.576	.336	.187	.448	.071	.037	.887

Productive contact	.113	.666	.213	.411	.400	.111	.037	.887
Multimedia contact	.223	.390	.186	.476	.457	.065	.075	.776
Reading contact	-.227	.381	.328	.198	.343	.178	-.075	.776

K4. Correlation between PV learner corpus and social network variables

Time 1

	Sophistication (n=17)		PV number (n=18)	
	ρ	p	ρ	p
L1 network size	.194	.471	-.146	.575
English network size	-.361	.186	-.157	.561
Network dispersion	.295	.328	-.210	.470
Topic multiplexity	-.158	.546	-.272	.274
Activity multiplexity	-.417	.096	-.146	.563
Network density	-.246	.376	-.188	.486
Network intensity	-.071	.802	-.171	.526
Friends' L2 proficiency	.466	.080	.015	.955
Face-to-face frequency	.141	.617	.099	.716
Online contact frequency	.092	.744	-.428	.098

	Semantic accuracy (n=17)		Grammatical accuracy (n=17)	
	ρ	p	ρ	p
L1 network size	-.082	.761	-.021	.938
English network size	.550*	.034	.011	.969
Network dispersion	.335	.263	.396	.180
Topic multiplexity	.524*	.031	-.089	.734
Activity multiplexity	.338	.184	-.064	.806
Network density	.353	.197	.197	.482
Network intensity	.214	.445	-.348	.204
Friends' L2 proficiency	.157	.577	.472	.076
Face-to-face frequency	-.110	.697	-.408	.131
Online contact frequency	-.051	.857	-.267	.336

Time 2

	Sophistication (n=14)		PV number (n=13)	
	ρ	p	ρ	p
L1 network size	.359	.208	.385	.194
English network size	.532	.075	-.111	.731
Network dispersion	-.200	.534	-.733**	.007
Topic multiplexity	-.215	.460	-.200	.513
Activity multiplexity	-.237	.414	.221	.468
Network density	.288	.365	-.219	.493
Network intensity	.575	.050	.351	.264
Friends' L2 proficiency	-.562	.057	-.017	.959
Face-to-face frequency	.007	.983	-.308	.330
Online contact frequency	.243	.446	-.136	.673

	Semantic accuracy (n=14)		Grammatical accuracy (n=14)	
	ρ	p	ρ	p
L1 network size	-.424	.131	-.167	.568
English network size	-.492	.104	-.216	.501
Network dispersion	-.398	.200	.240	.452
Activity multiplexity	.051	.863	.292	.311
Topic multiplexity	-.410	.146	.139	.635
Network density	-.173	.590	.044	.892
Network intensity	-.531	.075	-.453	.140
Friends' L2 proficiency	.609*	.036	-.426	.167
Face-to-face frequency	-.658*	.020	-.154	.633
Online contact frequency	.243	.446	-.136	.673

K5. Standard regressions models for Time 1 PV scores

T1 Hierarchical regression Model 1 results of productive PV score

Model 1	B	SE	Beta	t	Sig.
(Constant)	7.990	1.624		4.920	.000
Chinese nationality	.529	.687	.072	.769	.443
VLT1	.235	.039	.528	5.972	.000
motivation	-.422	.326	-.113	-1.293	.199
Receptive contact	.382	.362	.109	1.053	.295
Speaking contact	.076	.361	.021	.211	.833
Writing contact	-.065	.350	-.017	-.187	.852

T1 Standard regression Model 2 for receptive PV scores. Did not converge.
[F (3, 120) =1.847, $p=.142$, $r^2=.044$]

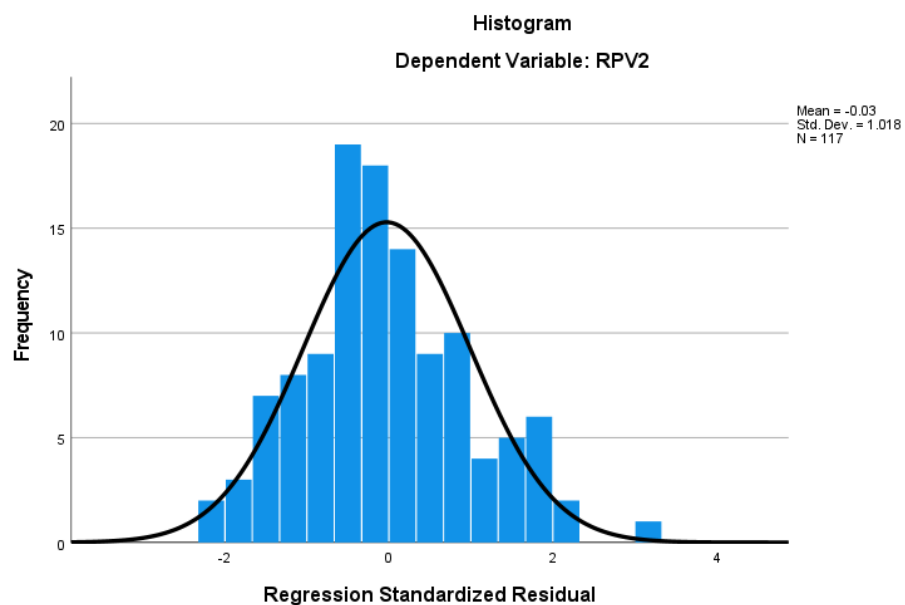
T1 Hierarchical regression Model 1 results of productive PV scores

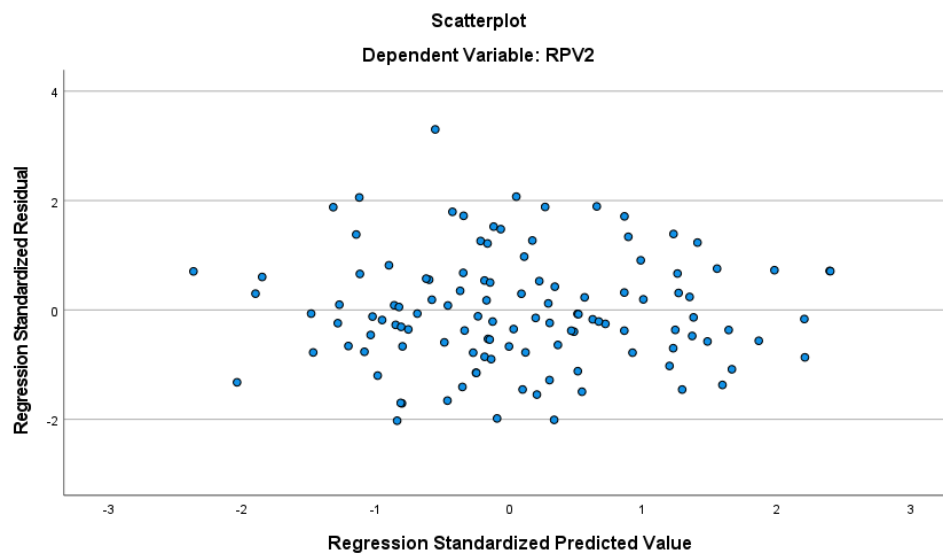
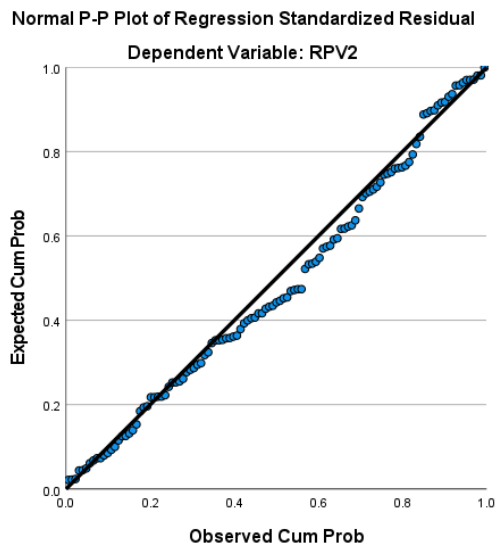
Model 1	B	SE	Beta	t	Sig.
(Constant)	4.437	1.488		2.981	.003
Chinese nationality	1.054	.630	.160	1.673	.097
VLT1	.183	.036	.461	5.071	.000
motivation	-.253	.299	-.076	-.848	.398
Receptive contact	.575	.332	.183	1.732	.086
Speaking contact	-.081	.331	-.025	-.243	.808
Writing contact	.096	.321	.028	.300	.764

T1 Standard Regression Model 2 for receptive PV scores. Did not converge
[F (3, 126) =2.562, $p=.058$, $r^2=.058$]

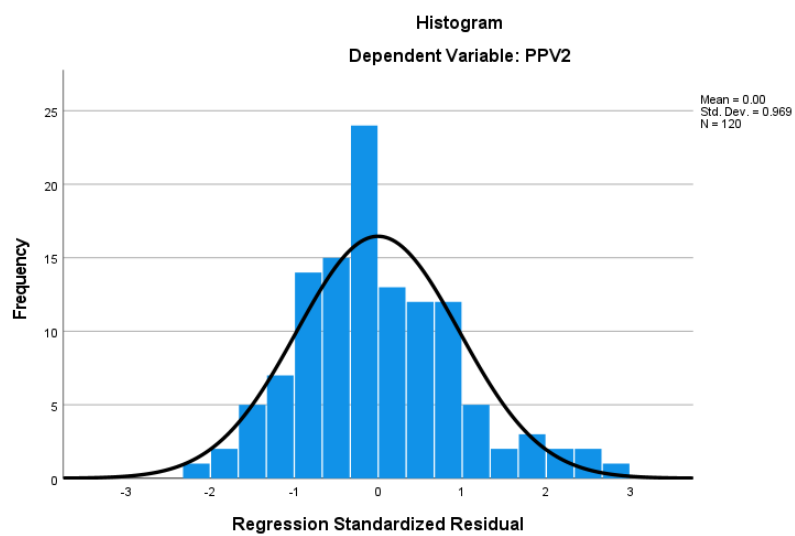
K6. Time 2 Hierarchical regression model residual checks:

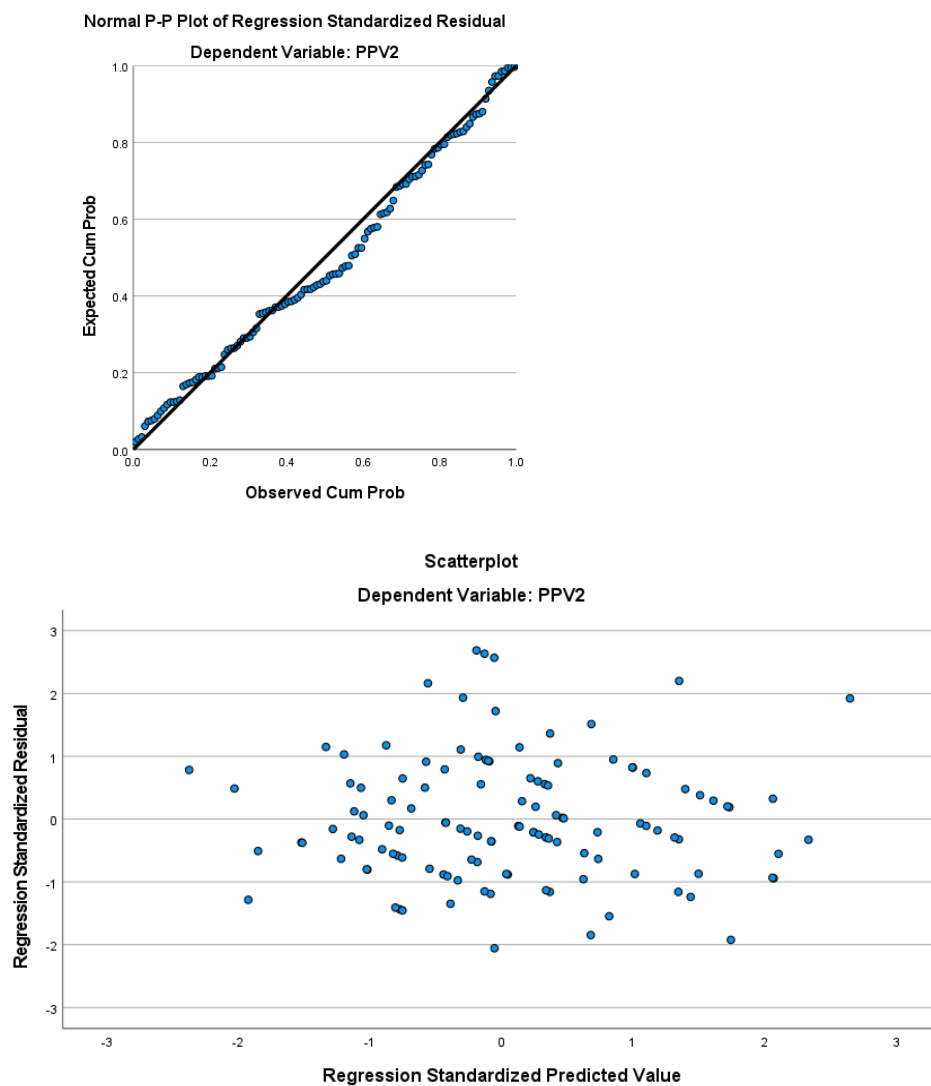
Model 1 Receptive PV scores





Model 1 Productive PV scores





T2 Standard Regression Model 2 for receptive PV scores

[F (3, 121) =12.462, $p < .001$, $r^2 = .236$]

Model 2	B	SE	Beta	t	Sig.
(Constant)	10.934	.299		36.608	.000
Multimedia contact	1.226	.501	.309	2.446	.016
Productive contact	-.966	.473	-.243	-2.042	.043
Reading contact	1.607	.451	.383	3.564	.001

T2 Standard Regression Model 2 for productive PV scores

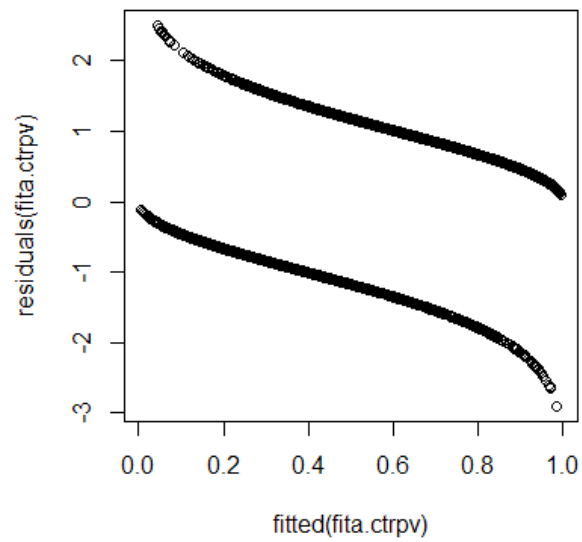
[F (3, 124) =12.031, $p < .001$, $r^2 = .225$]

Model 2	B	SE	Beta	t	Sig.
(Constant)	8.008	.244		32.764	.000
Multimedia contact	1.036	.410	.317	2.525	.013
Productive contact	-.852	.387	-.260	-2.202	.030
Reading contact	1.282	.369	.372	3.476	.001

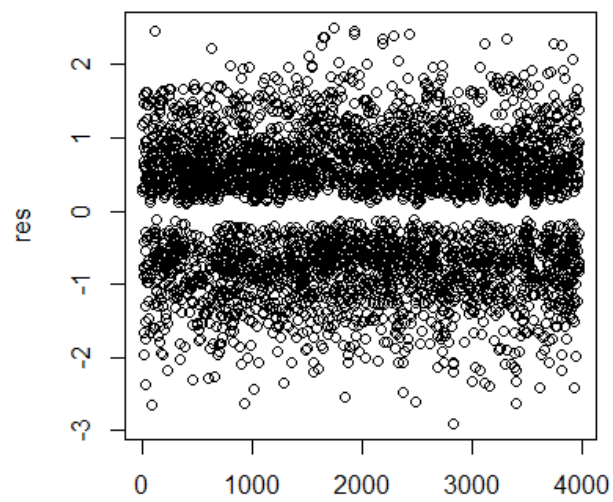
K7. Residuals checks of mixed-effects modeling

Receptive PV scores

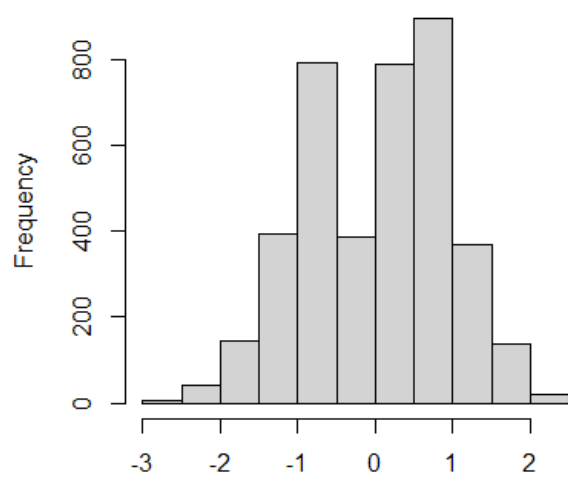
Residual plot



Scatterplot

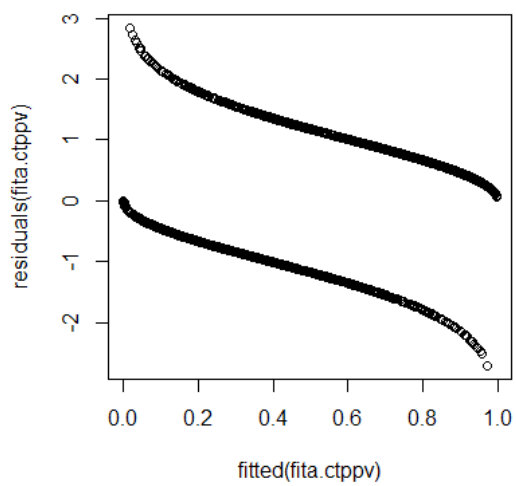


Histogram

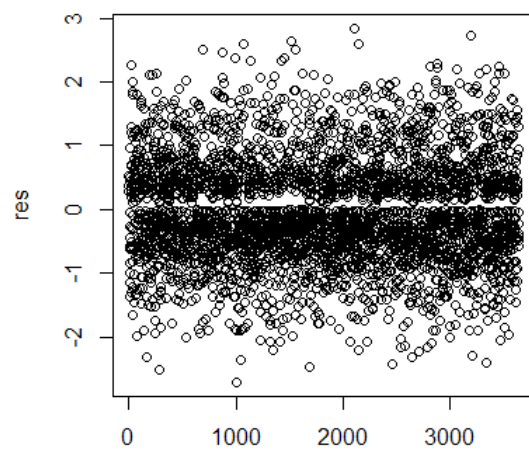


Productive PV scores

Residual plot



Scatterplot



Histogram

