

Durham E-Theses

*The Traveller Has Many a Tale to Tell: Personal
Change and Language Development as a Result of a
Year Abroad*

IAN JOHN MOORE

How to cite:

MOORE, IAN JOHN (2020) *The Traveller Has Many a Tale to Tell: Personal Change and Language Development as a Result of a Year Abroad*. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a <https://etheses.durham.ac.uk/id/eprint/13887/> is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

Abstract

Each year, hundreds of thousands of students worldwide undertake a study abroad experience. It is widely assumed that these experiences can be *transformative* and *life-changing* for learners but to date, little evidence exists which has empirically challenged these claims. The objective of this thesis is therefore to evaluate the efficacy of the UK year abroad experience regarding fostering positive personality changes, well-being and linguistic development. In doing so, the thesis looks to provide the reader with both a descriptive and explanatory account of change. To achieve this, the thesis comprises of two studies, which complement each other.

The first, Study 1, comprises of a systematic review of Second Language Acquisition literature within the domain of study abroad to ascertain the extent to which sojourning fosters language acquisition compared to remaining at-home. The findings indicate limited robust evidence from which to draw causal claims regarding the extent to which sojourning accounts for linguistic growth. Seven studies were carried forward to the in-depth review, suggesting sojourning to facilitate global proficiency and oral fluency, but where gains were found, these tended to be highly individualised. The review further supports the claims that growth experienced during a study abroad period is highly compartmentalised and subtle in nature.

The second, Study 2, collected data across the academic year 2018/19 in 180 students, 110 who studied abroad, and 70 who remained at-home in their final year of undergraduate study. Undertaking multilevel modelling and regression analysis, the findings indicated that sojourners became significantly more agreeable and curious while abroad compared to non-sojourners. Those who perceived themselves to be less lonely and felt part of the community appeared to experience the most value from the experience in relation to personal growth and mental well-being. An intensive repeated measurement component was also undertaken, demonstrating within-person variability to be systematically related to perceived situational antecedents.

Together, the two studies indicate that the year abroad experience is not *transformative* for all and that for some, the year abroad may pose distinct and unique challenges not faced by those in an at-home environment. The thesis offers stakeholders an insight into tools and actions which can be taken to aid student acculturation and ultimately foster positive personal and linguistic change. In light of the current political climate in the UK, this thesis comes at an important time and may hold relevance to policymakers and stakeholders alike as we enter into a *new* Europe.

**THE TRAVELLER HAS MANY A TALE
TO TELL:**

**PERSONAL CHANGE AND LANGUAGE
DEVELOPMENT AS A RESULT OF A
YEAR ABROAD**



IAN. J. MOORE

*St Mary's College
School of Education*

Durham University

*This dissertation is submitted for the degree of Doctor in Philosophy
September 2020*

Table of Contents

<i>List of Figures</i>	7
<i>List of Appendices</i>	7
<i>List of Tables</i>	8
<i>Acknowledgements</i>	10
<i>Introduction</i>	11
<i>Chapter 1: Residency Abroad</i>	12
1.1 Terminology	12
1.2 A Brief History of Residency Abroad and its Research.....	13
1.3 Study Abroad: A European Perspective	13
1.4 Residency Abroad: The Global Picture.....	16
1.5 Language Learners: The Specificity of The British Experience	16
1.6 Domestic Context: Life at a British University	17
1.7 The Future of Exchange Programmes in the UK	17
1.8 Recognising the Overarching Aims of the Thesis.....	18
<i>Chapter 2: Personality and Well-being</i>	19
2.1 Personality and its Make-up	19
2.1.1 Trait personality.....	19
2.1.2 State personality.....	19
2.2 Trait Personality Taxonomies.....	20
2.3 From Trait to State: A Change in Focus	21
2.4 The Power of the Situation in State Personality	23
2.4.1 Defining the situation	23
2.4.2 Cues, Characteristics and Classes	24
2.4.3 Providing a situational framework	24
2.5 Experience Sampling Methodology.....	25
2.6 Principles of Personality Development	26
2.7 Personality Development Between 18 – 25: The Age of Wonder.....	28
2.8 Personality and Life Events	29
2.9 The Role of Learning Context in Accounting for Personality Development.....	30
2.9.1 <i>Study abroad</i>	30
2.9.1.1 Self-selection effects	30
2.9.1.2 Socialization effects (broad personality traits).....	31
2.9.1.3 Socialization effects (narrow personality traits).....	32
2.9.2 <i>Undergraduate study</i>	34
2.9.2.1 Socialization effects (broad personality traits).....	34
2.9.2.2 Socialization effects (narrow personality traits).....	35
2.10 Personality, Study Abroad and its Differing Paradigms.....	36
2.11 Well-being and the Role of Learning Context in Education	38

2.11.1	Study abroad (sojourners)	39
2.11.2	At-home instruction (non-sojourners).....	41
Chapter 3:	Second Language Acquisition	42
3.1	Study Abroad: The ‘Golden Ticket’ for Linguistic Change?	42
3.2	Variables Impacting Linguistic Gain.....	45
3.3	Personality Within the Field of Second Language Acquisition.....	48
3.3.1	Second Language Acquisition and the broad traits	49
3.3.2	Second Language Acquisition and the narrow traits	51
3.4	Personality, Second Language Acquisition and Study Abroad	52
3.5	An Overview of the Thesis’s Research Questions	53
Chapter 4:	Study 1 – Value of Study Abroad on Linguistic Gain; a Systematic Review	56
4.1	Introduction.....	56
4.2	Design and Methodology	60
4.2.1	Research question	62
4.2.2	The importance of study design.....	62
4.2.3	Inclusion/Exclusion criteria.....	65
4.2.4	Study designs	66
4.2.4.1	Non-ERASMUS.....	67
4.2.4.2	ERASMUS	68
4.2.5	Electronic searching.....	69
4.2.6	Screening and data extraction process.....	71
4.2.6.1	Abstract screening	71
4.2.6.2	Screening of full papers.....	71
4.2.6.3	Data extraction	71
4.2.6.4	Mapping synthesis	72
4.2.6.5	In-depth review synthesis.....	72
4.3	Results	74
4.3.1	Searching procedure.....	74
4.3.1.1	Initial search	74
4.3.1.2	1 st stage screening (screening on title and abstract)	74
4.3.1.3	2 nd stage screening (screening on full text).....	74
4.3.1.4	3 rd stage screening (data extraction).....	75
4.3.1.5	Quality assurance.....	75
4.3.2	Study characteristics of all included studies	77
4.3.2.1	Synthesis of included studies by intervention	80
4.3.2.2	Summary of mapping table	82
4.3.2.3	Narrative synthesis: the in-depth review	85
4.3.2.4	The methodological quality of in-depth review studies	91
4.3.3	Results of the in-depth review	97
4.3.3.1	Oral.....	97
4.3.3.2	Writing	99
4.3.3.3	Reading	100
4.3.3.4	Grammar.....	101
4.3.3.5	Vocabulary (Lexical Richness).....	102
4.3.3.6	General Proficiency	103
4.4	Discussion	104
4.4.1	Linguistic outcome under study.....	104
4.4.2	Length of stay	105
4.4.3	Policy and practice	106
4.4.4	Strengths and limitations	106

4.4.5	Future directions	107
4.4.6	Conclusions and recommendations	107
Chapter 5: Study Two – Longitudinal Study.....		109
5.1	Introduction.....	109
5.1.1	Background context of the study.....	109
5.1.2	Aims and objectives.....	111
5.1.3	Rationale	111
5.2	Methodology.....	114
5.2.1	Sampling and participants	114
5.2.1.1	Sampling strategy	114
5.2.1.2	Recruitment.....	115
5.2.1.3	Response rate	117
5.2.1.4	Sample characteristics	118
5.2.1.5	Minimising missing data.....	122
5.2.2	Materials and data collection	123
5.2.2.1	Broad personality trait questionnaire	123
5.2.2.2	Narrow trait questionnaire	124
5.2.2.3	Well-being questionnaire.....	125
5.2.2.4	Experience Sampling Questionnaire.....	125
5.2.2.5	Linguistic proficiency.....	126
5.2.2.6	Focus groups	130
5.2.3	Study design.....	130
5.2.3.1	Personality and well-being.....	130
5.2.3.2	Language	132
5.2.4	Procedure	133
5.2.5	Piloting	134
5.2.6	Ethical considerations.....	135
5.2.7	Data analysis	137
5.2.7.1	Multiplicative effect (time * learning context)	137
5.2.7.2	Capturing the main effect of time	139
5.2.7.3	Predictors of outcome change and performance	139
5.2.7.4	State personality.....	141
5.2.7.5	Language change	142
5.2.7.6	Qualitative analysis	142
5.3	Results	143
5.3.1	Preliminary analysis of personality trait data.....	143
5.3.1.1	Summary of findings.....	146
5.3.2	Understanding the role of learning context in broad and narrow trait change	147
5.3.2.1	Capturing differing narrow trait trajectories between learning contexts.....	152
5.3.2.2	Summary of findings.....	157
5.3.3	Assessing the main effect of time in broad trait trajectories.....	158
5.3.3.1	Qualitative analysis	160
5.3.3.2	Summary of findings.....	165
5.3.4	Predictors of individual differences in broad trait trajectories	166
5.3.4.1	Change in trait openness	168
5.3.4.2	Change in trait conscientiousness.....	171
5.3.4.3	Change in trait extraversion	175
5.3.4.4	Change in trait agreeableness	177
5.3.4.5	Change in trait neuroticism	179
5.3.4.6	Qualitative findings.....	181
5.3.4.7	Summary of findings.....	182
5.3.5	State personality and situational contingencies	183
5.3.5.1	Descriptive statistics	183
5.3.5.2	The relationship between the broad traits and states.....	184

5.3.5.3	Capturing variability across the personality levels	185
5.3.5.4	Systematicity of variability in state agreeableness	187
5.3.5.5	Summary of findings.....	190
5.3.6	Capturing well-being change and its predictors.....	191
5.3.6.1	Understanding the role of context on psychological well-being	192
5.3.6.2	Well-being change over time.....	193
5.3.6.3	Predictors of fluctuations in psychological well-being.....	195
5.3.6.4	Predictors of average psychological well-being across the year.....	197
5.3.6.5	Qualitative findings.....	199
5.3.6.6	Summary of findings.....	202
5.3.7	Language change and its predictors	203
5.3.7.1	Observed Language change in learners across the year	203
5.3.7.2	Perceived language changes.....	206
5.3.7.3	Qualitative findings.....	206
5.3.7.4	Predictors of language change.....	208
5.3.7.5	Summary of findings.....	211
5.4	Discussion and conclusions.....	212
Chapter 6: Discussion and Conclusions		216
6.1	Returning to the Purpose of the Thesis	216
6.2	Does the Year Abroad Work? Evaluating the Value of a Sojourn Experience.....	218
6.2.1	Evaluating the value of sojourning on personal growth (Study 2)	218
6.2.2	Evaluating the value of sojourning on psychological well-being (Study 2).....	224
6.2.3	Evaluating the value of sojourning on language learning (Study 1 & Study 2)	226
6.2.4	Evaluating the experience as a whole	232
6.3	Implications.....	233
6.3.1	Implication for theoretical development.....	233
6.3.1.1	The year abroad as a valid life event?.....	233
6.3.1.2	Principle of personality development	233
6.3.1.3	Capturing variability and ascertaining its systematicity	234
6.3.2	Implications for policy and practitioners	235
6.3.2.1	The mid-year dip	236
6.3.2.2	Reducing loneliness and promoting extracurricular participation	236
6.3.2.3	Challenging the notions of linguistic ‘nativeness’	237
6.4	Strengths and Limitations.....	237
6.4.1	Strengths	237
6.4.1.1	Sample make-up	237
6.4.1.2	Providing tentative causal conclusions regarding the value of sojourning.....	238
6.4.2	Limitations	238
6.4.2.1	Sample size.....	238
6.4.2.2	Threat of potential biases.....	239
6.5	Future Directions.....	241
6.5.1	Exploring the differences between compulsory and voluntary sojourners	241
6.5.2	Ensuring language studies hold methodological rigour.....	241
6.5.3	Re-entry and long-term personality change.....	242
6.6	Conclusions	242
References		244

List of Figures

Figure 1: Web of Science Search Strategy	70
Figure 2: PRISMA Flow-Diagram	76
Figure 3: Overview of Data Collection Sequence	133
Figure 4: Broad Trait Interaction Plots Between Time and Learning Context.....	149
Figure 5: Narrow Trait Interaction Plots Between Time and Learning Context.....	154
Figure 6: Broad Trait Personality Trait Change Histograms (T4 – T1).....	160
Figure 7: Conscientiousness Change Interaction Effect (Employment Status * Time) in Sojourner Sample	174
Figure 8: Between and Within-Person Variability in Measured Personality Variables (across all learning contexts)	186
Figure 9: Interaction Plot of Relationship Between State Agreeableness and State Duty across Learning Contexts	188
Figure 10: Overview of Average Well-being Scores across Time and Learning Context	191
Figure 11: Panel Plots of Well-being Scores across Time.....	193
Figure 12: Average Proficiency Scores for Language Learners across Time	204
Figure 13: Language Change Interaction Effect (Time * Honours Programme)	205

List of Appendices

Appendix A: Systematic Review Protocol.....	269
Appendix B: Systematic Review Search Strings	277
Appendix C: Data Extraction Tool	281
Appendix D: Systematic Review Mapping Table (40 studies)	293
Appendix E: PRISMA Checklist	304
Appendix F: Demographic Questionnaire	306
Appendix G: Broad Trait Personality Questionnaire	313
Appendix H: Narrow Personality Trait Questionnaire.....	316
Appendix I: State Personality Questionnaire	325
Appendix J: Language Instruments	331
Appendix K: Perceived Competencies Questionnaire.....	350
Appendix L: Links to Language Instruments.....	352
Appendix M: Language Instrument Pilot Statistics	353
Appendix N: Focus Group Transcription Example.....	355
Appendix O: Participant Sheet (Study 2)	370
Appendix P: Study 2 Ethical Approval.....	372
Appendix Q: Participant Information Leaflet (Focus Groups)	373
Appendix R: Focus Groups Consent Form.....	374
Appendix S: Focus Group Ethical Approval	375

List of Tables

Table 1: The Five-Factor Model Descriptors (taken from Ożańska-Ponikwia & Dewaele, 2012, p. 118).....	20
Table 2: Principles of Personality Development (Roberts & Nickel (2017, p. 162).....	28
Table 3: Inclusion/Exclusion Criteria.....	66
Table 4: Inclusion/Exclusion Study Design Criteria (non-ERASMUS)	68
Table 5: Inclusion/Exclusion Study Design Criteria (ERASMUS).....	68
Table 6: Electronic Search Results Prior to Deduplication.....	74
Table 7: Frequency of Included Studies according to Study Design.....	77
Table 8: Frequency of Included Studies according to Study Outcome	77
Table 9: Frequency of Included Studies according to Publication Date	78
Table 10: Frequency of Included Studies according to Sample Size	78
Table 11: Frequency of Included Studies according to Target Language.....	79
Table 12: Frequency of Included Studies according to Length of Stay.....	79
Table 13: Study Designs (In-depth Review).....	86
Table 14: Study Outcomes (In-depth Review)	86
Table 15: Formation of the Treatment Group (In-depth Review).....	86
Table 16: Formation of the Control Group (In-depth Review).....	86
Table 17: Mapping Table (In-depth Review)	87
Table 18: Methodological Quality of Included In-depth Studies	89
Table 19: Synthesis Without Meta-analysis (SWiM) Items	96
Table 20: Overview of Universities Selected.....	115
Table 21: Overview of Instrument Response Rate.....	118
Table 22: Overview of Sample Demographic Characteristics.....	119
Table 23: Overview of Host Countries	120
Table 24: Departure Month (Start of the Year Abroad).....	121
Table 25: Returning Month (End of the Year Abroad).....	121
Table 26: Sojourner Roles and Accommodation.....	121
Table 27: Overview of Focus Group Participants.....	122
Table 28: Overview of Study Outcomes.....	128
Table 29: Overview of Programme-specific Variables	128
Table 30: Overview of Predictors Associated with Environmental Interpretation	129
Table 31: Overview of Variables captured in the Experience Sampling Questionnaire	129
Table 32: Means and SDs of Baseline Personality according to Drop-out Condition.....	143
Table 33: Logistic Regression Pertaining to Broad Traits as Predictors of Drop-out.....	144
Table 34: Means and SDs of Baseline Personality Traits (Voluntary Sojourners vs At-home Learners).....	145
Table 35: Means and SDs of Baseline Personality Traits (Voluntary vs Compulsory Sojourners).....	146
Table 36: Means and SDs of Broad Personality Traits across Time according to Learning Context.....	148
Table 37: Parameter Estimates for Linear Growth Model of Broad Traits as a Function of Learning Context	150
Table 38: Means and SDs of Narrow Personality Traits across Time according to Learning Context.....	153
Table 39: Parameter Estimates for Linear Growth Model of Broad Traits as a Function of Learning Context	155

Table 40: Parameter Estimates for Linear Growth Model of Broad Traits across Time (Sojourners)	159
Table 41: Descriptive Statistics for Broad Trait Slopes Estimated by MLM (Sojourners).....	159
Table 42: Testing Potential Covariates of Broad Trait Change from T1 to T4 (Sojourners).....	167
Table 43: Multiple Regression of Change in Openness on the Identified Covariates (Sojourners)	168
Table 44: Multiple Regression of Change in Conscientiousness on the Identified Covariates (Sojourners)	172
Table 45: Multiple Regression of Change in Extraversion on the Identified Covariates (Sojourners)	175
Table 46: Multiple Regression of Change in Agreeableness on the Identified Covariates (Sojourners)	178
Table 47: Multiple Regression of Change in Neuroticism on the Identified Covariates (Sojourners)	179
Table 48: Descriptive State Personality Statistics (Means/SDs)	183
Table 49: Parameter Estimates for Linear Growth Model of State Personality as a Function of Broad Traits	184
Table 50: Between and Within-person Variability at each Measured Personality Level.....	185
Table 51: Parameter Estimates for Linear Growth Model of State Personality as a Function of Situation.....	187
Table 52: Descriptive Statistics for Situation-Contingent Agreeableness Slopes Estimated by MLM.....	190
Table 53: Parameter Estimates for Linear Growth Model of Well-being as a Function of Learning Context	192
Table 54: Parameter Estimates for Linear Growth Model of Well-being as a Function of Time	194
Table 55: Monthly Well-being Scores as a Function of Monthly Narrow Trait Scores (sojourners)	195
Table 56: Multiple Regression of Average Well-being Scores on Baseline Broad Traits	197
Table 57: Multiple Regression of Average Well-being Scores on Environmental Variables.....	198
Table 58: Simple Main Effect for Group on Language Change	205
Table 59: Simple Main Effect for Time on Language Change.....	205
Table 60: Perceived Linguistic Ratings (Means/SDs).....	206
Table 61: Testing Potential Covariates of Language Change from T1 to T3	210
Table 62: Multiple Regression of Identified Covariates in L2 Proficiency Change (Baseline Openness).....	210

Statement of Declaration

I declare that this work is my own. Material in this thesis has been submitted for publication which is currently in review.

Statement of Copyright

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

I wish to use this space to take the opportunity to say thank you for all the people who have helped me to complete my doctoral study.

First and foremost, I would like to express my gratitude to my supervisors, Dr Nadin Beckmann, Prof Carole Torgerson and Dr Bingham Zheng. Their guidance has been invaluable in shaping this piece of work, and I will be forever thankful for the feedback they have provided.

My gratitude must also extend to my close friends, family and partner, Natalia, whose support and patience during the doctoral study has been immeasurable.

I would, of course, also like to extend my gratitude to all the participants who partook in the study. Without them, this research project simply wouldn't have been possible. Similarly, I would like to say thank you for all the administrative staff who helped me find the participants in the first instance.

Lastly, I would like to express my gratitude to the ESRC, who through their scholarship, I was able to complete this doctoral study.

Introduction

An ever-increasing number of students worldwide take up the opportunity to complete an exchange programme each year. These programmes serve as a natural hiatus from their academic studies with the purpose of immersion into a new culture or language of a foreign country. The push by stakeholders to support and promote these experiences has today meant that learners have greater opportunities than ever before to gain valuable experience abroad.

Within the British context, in which this thesis is set, sojourners are provided with two types of exchange programmes. The first, and most common, is run under the European Community Action Scheme for the Mobility of University Students programme (henceforth ERASMUS), which, since its inauguration in 1987, enabled academic staff and students alike to exchange with European chartered institutions or work for affiliated organisations (e.g., the British Council) and private businesses. The programme makes up part of the larger European Union's drive to facilitate mobility and multilingualism within its citizens (Teichler, 2015). In 2015-2016, 303,880 students partook in the ERASMUS+ programme, making it the largest exchange programme in the world (European Commission, 2019). The second type is through means of an overseas exchange, which allows learners of all disciplines to work and study in almost any part of the world through private agreements created by tertiary agreements.

This thesis takes a cross-disciplinary approach, exploring constructs found in both the fields of Psychology and Applied Linguistics. Its purpose, in brief, is to capture the value of sojourning in relation to facilitating positive changes in the area of personality, well-being and language change. Using a longitudinal design, the thesis looks to objectify the changes so often highlighted in research which sees the year abroad as a *transformative* experience for all, and which typically use qualitative approaches to ascertain perceived change retrospectively (e.g., McLeod & Wainwright, 2009). As Byram (2008) notes, it is these claims which must continue to be empirically challenged in order to ensure stakeholders are appropriately informed.

The thesis is split into six chapters. Chapters one, two and three introduce readers to the constructs which the thesis investigates. Chapter four presents the systematic review, while chapter five presents the results of the empirical study. The concluding chapter offers an insight into the contribution this study makes in the field, together with limitations and directions for future educational and research practice.

Chapter 1: Residency Abroad

The following chapter aims to align readers with the concept of residency abroad. Conceptual issues are first introduced, followed by a historical overview of global exchange programmes. An explanation of the different types of exchange programmes, together with a description of the sojourn experience specific to UK-domiciled students, is then given. Lastly, the current political perspective is explored regarding the future of sojourning programmes in the UK, followed by an overview of the aims and objectives of the thesis.

1.1 Terminology

Within the international mobility literature, the concept of sojourning has been met with a plethora of definitions and conceptions (Jackson, 2018). The most frequently used term is *study abroad*, defined by Kinginger (2009, p. 11) as “a temporary sojourn of pre-defined duration, undertaken for educational purposes.” Within the British context, however, scholars have questioned whether *study abroad* adequately captures the specificities of the experience for learners. Coleman (1998), for example, argues against using *study abroad* for its concerns only one type of sojourner role, that of studying. Coleman (1998, p. 174) instead prefers the term *residency abroad* for as he notes, it is the “most widely accepted term, embracing the three most popular ways of spending the period abroad: as a foreign language assistant, on a work placement, or as a student.” Alred and Byram (2002) further argue that using *study abroad* solely portrays the experience to be voluntary, which for many British undergraduates is untrue. Depicting both the length and compulsory nature of a sojourn, the term *year abroad* is quintessentially British (Alred & Byram, 2002). In light of this discussion, this thesis will use the terms *study abroad*, *residency abroad* and *year abroad* interchangeably to denote a temporary sojourn in a foreign country.

The thesis also refers to the term personal *growth* and personality *change*. While these terms may appear ambiguous at first, their usage is rather specific. The former refers to a shift in personality which may be considered a positive shift in behaviours and habits reflective of increased maturity. For example, an individual becoming more conscientious and less neurotic would be representative of personal *growth*. The latter on the other hand is broader in nature and refers to a shift in personality which is either an increase or decrease in trait-like cognitive, affective, and/or behavioural responses. For example, an individual can become more open or less open and that would be considered personality *change*.

1.2 A Brief History of Residency Abroad and its Research

Each year thousands of students worldwide undertake an overseas exchange programme. These experiences are hypothesised to allow for immersion into a new culture and/or language in a foreign country. While for many, the notion of international mobility may be a relatively recent phenomenon, its roots are almost a thousand years old. Records show that the first recorded sojourner was Emo of Friesland who in 1190, left his hometown in northern Holland to begin a period of study at Oxford University (Lee, 2015). In the 700 years that followed, war, poverty, and poor educational policy often meant travel was reserved for the rich and elite. However, by the turn of the 20th century, the foundations had been laid for change to occur and although the onset of the two Great World Wars temporarily stagnated progress, their conclusions sparked a widespread call for government and policymakers to support cultural exchange programmes. Global leaders such as President Roosevelt saw these programmes as a tool to facilitate intercultural communication and bring with-it peace. By the end of the 1980s and with the raising of the Iron Curtain, many of the geopolitical obstacles which had once hampered international mobility were removed (Lee, 2015). Driven by globalisation and technological advances, it is perhaps in the past two decades in which sojourning has seen its most radical transformation (Donatelli, 2010). The advent of English as a Lingua Franca and the ability to stay in contact with friends and family at-home has led many scholars to question whether a period abroad can be described as *immersive* (e.g., Coleman & Chafer, 2010).

There is today interest in a wide variety of variables associated with the *study abroad experience*. These include Second Language Acquisition (e.g., Segalowitz & Freed, 2004; Serrano, Tragant & Llanes, 2012); learner identity (e.g., Gieser, 2015; Sigalas, 2010), personality (e.g., Niehoff, Petersdotter, & Freund, 2017; Zimmermann & Neyer, 2013), employability (e.g., Engel, 2010; Teichler & Janson, 2007) and intercultural sensitivity (e.g., Beaven & Spencer-Oatey, 2016; Deardorff, 2006). Research is required to ensure that the resources and efforts expended by stakeholders are worthwhile and of benefit (Sanz & Morales-Front, 2018).

1.3 Study Abroad: A European Perspective

Driven by a desire to increase multilingualism and citizen mobility within the European Union (henceforth EU), higher education policy has long supported exchange programmes (Teichler, 2015). ERASMUS (+) is today the result of this support, inaugurated into the European Community in 1987. Upon its introduction, the programme had the following aims and objectives as set out in its founding document (Council of Ministers, 1987, pp. 21-22):

“The objective of the ERASMUS programme shall be (...)

(i) that the Community may draw upon an adequate pool of manpower with first-hand experience of economic and social aspects of other Member States (...)

(iv) to strengthen the interaction between citizens in different Member States with a view to consolidating the concept of a People’s Europe;

(v) to ensure the development of a pool of graduates with direct experience of intra-Community cooperation, thereby creating the basis upon which intensified cooperation in 28 the economic and social sectors can develop at Community level.”

The programme is today considered a huge success for the EU (Wodak, 2017) and has so far supported over three million people to work, study and train abroad, making it the largest exchange programme in the world. In its original capacity, the ERASMUS programme enabled mobility within European borders only. However, with the introduction of the ERASMUS+ programme in 2014, opportunities became open to a non-European domiciled population. The programme is currently made up of 33 programme countries (28 EU member states and five other European countries) and partner countries located all over the world, including Canada and the USA. The programme enables not only university students and staff to travel abroad but also trainees, apprentices, volunteers, youth workers and other professionals, tackling many criticisms ERASMUS has faced regarding elitism (Wodak, 2017).

In its latest report, ERASMUS+ enabled 312,347 university students into higher educational mobility in the academic year 2015/2016 (European Commission, 2019). Spain (47,138), Germany (32,876), France (29,068), Italy (25,108) and the UK (31,243) make up the top five receiving countries, while France (43,769), Germany (40,629), Spain (39,759), Italy (35,371) and Turkey (16,889) make up the top five sending country.

For many students in Europe and beyond, partaking in ERASMUS is a highly competitive process. Fierce selection criteria (e.g., proficiency, personal statements, interviews) often means there are just as many rejections as acceptances, regardless of learner intention or aspiration to sojourn (Hessel, 2017). For British-domiciled students, however, the process is relatively less competitive. British institutions are paired with other European universities according to their European Charter status, ensuring the institution is accredited according to the quality standards set by the EU. Those who wish to study select a partnered institution, with much of the paperwork completed at an administrative level. For those on teaching placements, individuals

determine their destination according to chosen criteria (e.g., urban or rural; town or city), and while the process is more complex and competitive, the majority receive their first choice. For those working, discussions are more often conducted between the company and individual privately, with universities having little involvement in the administration process.

For British-domiciled learners, ERASMUS offer three types of sojourner-roles. Participants can either teach, study or work, or undertake a combination of roles, split across the year. Teaching programmes were first introduced in France in 1904 (British Council, 2005), and due to it being a paid position was the most popular option until ERASMUS funded students to study at university in the 1980s (Teichler, 2004). A decade later, studying had become the most popular option, with Coleman (1996) finding over half of the 12,000 British sojourners opting to study while abroad, with this popularity remaining today. Work placements always have and continue to be the least common of the three options (Mitchell, Tracy-Ventura, & McManus, 2017a).

Regarding the length of stay, learners generally spend between three and 12 months abroad (6.2 months on average studying; 4.4 months for work experience) according to the European Commission (2015). This period is considerably longer than the average length of stay for North American sojourners, of who typically remain abroad for an average of six weeks (Kinginger, 2015), reflecting the popularity found in North America for short-term exchange programmes.

Concerning accommodation, to the best of my knowledge, no published data exist regarding accommodation choice. Nonetheless, those who are studying tend to stay in either rented or university-owned accommodation with other exchange students, while those who work or teach tend to live either alone or in homestays (Mitchell et al., 2017a).

Support available to ERASMUS participants before and during their time abroad is very much dependent on the individual institution. Financially, support occurs through means of a grant, paid over two instalments, of which the vast majority (97%) of ERASMUS participants receive (Souto-Otero et al., 2013). This grant goes towards aiding accommodation and living costs and is beyond doubt one of the primary reasons why many can partake in the programme. Pre-departure workshops have been criticised with many students considering them inadequate (Jacobone & Mora, 2015). While abroad, this level of support is dependent on sojourner role. For example, while those studying may have fresher events, those teaching or working receive few introductory events and may find it challenging to meet and integrate with people their age.

Home institutions rarely instigate face-to-face contact with a counsellor, but nowadays many do offer access to online counselling services and more lecturers are offering office hours over Skype. Such measures can serve as essential support structures to sojourners and can go some way to improving their mental health (UMHAN, 2013).

1.4 Residency Abroad: The Global Picture

While Europe remains a top destination for many potential sojourners, international mobility beyond Europe is increasing. Such programmes are typically one academic year in length and are available in most departments (based on UCAS searches in preparation for Study 2). Typically, places on an overseas exchange are not guaranteed, even if the degree programme contains a year abroad component and acceptance on an overseas exchange programme depends on previous academic performance, application quality, and whether it is a partner university.

From a British perspective, although increasing, the number of learners undertaking international mobility is low. In the latest statistics available, 18,510 UK-domiciled students went abroad (including through ERASMUS+) in 2016/17, representing 7.8% of all graduating UK students (UUKI, 2019). This percentage is lower than the target set by the UK's Strategy for Outward Student Mobility who by 2020, hope to have 13.2% of UK-domiciled full-time students studying abroad as part of their first degree. In 2016/17, the disciplines of modern foreign languages (87.1%), medicine and dentistry (30.8%) and veterinary science (17.2%) saw the highest percentage of outgoing students according to the UUKI (2019). Over half (50.8%) of mobility activities were in Europe, with 18.5% in North America and 12.3% in Asia. More female students (11,270) went abroad from the UK in 2016/17 than males (7,235) (UUKI, 2019).

1.5 Language Learners: The Specificity of The British Experience

For British-domiciled students undertaking a modern foreign language degree, the year abroad serves as a compulsory component to their academic degree, taking place in their 3rd year of study. As of writing, only York St John and Sheffield Hallam offer a different degree pathway. While those at York St John complete their year abroad in the 2nd year of study, Sheffield Hallam allows their students to spend up to 18 months abroad; six months in the 2nd year and the entire 3rd year. Those who study two languages (Joint Honours) will most likely spend a semester in each country of their learnt languages. Those who learn one language (Single Honours) will either stay in the same country or live in two countries which speak the same language natively. For the few who study three languages, learners spend the summers in the third country. This

situation is relatively peculiar to the British context for in most European countries, sojourning, regardless of discipline, is a voluntary experience. Where it is compulsory, such as Spain (e.g., Pérez-Vidal, 2015), this period tends to be for one semester only.

1.6 Domestic Context: Life at a British University

For those who continue to study at a domestic institution, the 3rd year is often considered the most important because exam scores hold the most weighting. Academic pressure and job searching can make this a time of great stress for students (Stamp et al., 2015) with exams occurring after Christmas and at the end of May. Students will likely stay in private accommodation within settled friendship groups, but in most instances will still be considered an *outsider* within the local community outside of the university.

1.7 The Future of Exchange Programmes in the UK

On the 23rd of June 2016, The United Kingdom (henceforth the UK) voted to leave the European Union. Subsequently, questions were quickly raised concerning the participation of the UK in the ERASMUS programme post-2020. The UK currently finds itself in a period of transition, whereby the UK government are liaising and working with the European Union to confirm its status within the programme. The British government have not yet formally decided whether participation in the programme will continue and have so far voted against a clause which would have ensured that full membership in the programme continues. It has been confirmed that the UK will not be participating in the European Solidarity Corps programme; a programme designed to provide young people with opportunities to work or volunteer on domestic or international projects. According to ERASMUS+ (2020), the government is considering all possible alternatives to future international sojourn programmes, including a domestic alternative. While a domestic alternative may better reflect the government interests, there are questions surrounding whether funding would be available on the scale currently available through the European Union itself. For example, the European Commission in 2018 proposed doubling the amount of funding given to the ERASMUS programme to €30 billion. According to a report published by the House of Lords (2019), a domestic alternative would in no way come close to the benefits and monetary value offered through ERASMUS. The report further argued that the loss of the ERASMUS programme would “disproportionately affect people from disadvantaged backgrounds and those with medical needs or disabilities” (House of Lords, 2019, p.5). The UK government has now announced its own scheme to ensure study abroad can continue from September 2021 entitled the Turing scheme.

1.8 Recognising the Overarching Aims of the Thesis

Given the time, effort and resources expended towards study abroad programmes by both stakeholders and the students themselves, the overarching goal of the thesis is to evaluate the value of the sojourning experience across several important constructs. The thesis focuses on personal growth, linguistic growth and that of psychological well-being, all-important constructs in the field of study abroad research. The overarching research question can be defined as:

How effective is the year abroad experience in facilitating positive changes in the domains of personality, psychological well-being and linguistic competencies?

Given that the nature of the research question is evaluative, its purpose is to attribute any differences found between the treatment group (i.e., abroad) and the comparison group (i.e., at-home) to the intervention itself (i.e., study abroad). Given the compulsory nature of the year abroad in the UK for language learners, it was not possible to form a comparison group when evaluating changes in linguistic proficiency. Consequently, this thesis is divided into two studies.

Study 1 served as a systematic review which looked to synthesise the current literature regarding the extent to which sojourning fosters linguistic growth. By including only those studies which utilised a comparison group, and which were equivalent at baseline, the conclusions warranted causal inferences regarding sojourning having a direct effect on the observed linguistic change.

Study 2 collected primary data on a range of constructs. However, given that comparative data were only collected for the constructs of personality and psychological well-being, Study 2 cannot infer whether sojourning fosters linguistic proficiency. Study 2 provides not only a descriptive account of this change (i.e., *how*) but also looks to capture an explanatory account of this change (i.e., *why*), by capturing data on both personal, environmental and programme characteristics. Moreover, Study 2 comprises of an intensive repeated measurement of state personality, to ascertain the breakdown between within and between-person variability and whether this variability is systematic on situational characteristics.

Together, the two studies serve a complementary purpose, and space will be given in chapter six to discuss how these two studies have come together. Chapter two focuses on the first of the constructs investigated in this thesis, that of personality and well-being.

Chapter 2: Personality and Well-being

2.1 Personality and its Make-up

Personality can be defined as “the dynamic organisation within the individual of those psychophysical systems that determine his characteristic behaviour and thought” (Allport, 1961, p.28). It influences our behaviours and emotions and drives each individual to be unique. As a construct, personality can be viewed within the prism of traits and states, as explored below.

2.1.1 Trait personality

Trait personality refers to “relatively enduring patterns of thoughts, feelings, and behaviours that distinguish individuals from one another” (Bleidorn, Hopwood, & Lucas, 2018, p. 83). These traits serve as the building blocks of personality, each differing in weight, but which, when combined, results in an individual’s overall personality (Funder, 2006). While this weighting is considered relatively stable, there is evidence that substantial change can manifest itself across a person’s lifetime (e.g., Roberts, Walton, & Viechtbauer, 2006a). Traits are inheritable (Yamagata et al., 2006) and found to be universal across language and culture (McCrae & Terracciano, 2005). Variability in trait personality has been shown to be accountable by genetics with genes influencing up to 50% of inter-human differences in trait personality (Bouchard & McGue, 2003), while the environment accounts for the remaining variability.

2.1.2 State personality

State personality refers to the “short-term, concrete patterns of acting, feeling, and thinking” (Heller, Komar, & Lee, 2007, p. 899). These momentary patterns of behaviours reflect how the same individual can behave differently on different occasions. Despite their temporal differences, state personality can be measured through the same constructs and similar instruments as trait personality (Fleeson, 2001; 2007).

Conceptually, the two differ in that states require immediate recognition of current thoughts, feelings, and behaviours, while traits require the recognition of thoughts, feelings and behaviours *in general* over an extended period and are a representation of how the individual construes their self-identity (Beckmann & Wood, 2020).

2.2 Trait Personality Taxonomies

A major achievement of the past century in psychology, according to Dreary (2009), has been the development of the trait approach, which sees personality traits as hierarchically organised. At its apex stand (typically) five broad orthogonal dimensions (higher-order traits) which can then be divided into more narrow, specific facets known as lower-order traits (Dewaele, 2012). Amongst trait theorists, it is these higher-order traits which serve as the basis of one's personality and have the greatest power in describing personality (Eysenck, 1992; Funder, 2001). Moreover, it is these broad dimensions which are replicable across samples, inheritable and have a predictive capacity. While scholars (e.g., Ashton et al., 2004) have debated how many traits should make up the higher-order level, the most widely accepted number is that of five as conceptualised in the psychometrically derived Five-Factor Model (Costa & McCrae, 1985).

The Five-Factor Model is today the most widely cited model of personality structure (Soto, Kronauer, & Liang, 2016). According to the model, the common variance between primary personality variables can be explained by five broad, independent, and hierarchically organised dimensions (Salgado et al., 2015). These dimensions are as follows: openness, conscientious, extraversion, agreeableness and neuroticism. Table 1 outlines the characteristics of each trait.

Table 1: The Five-Factor Model Descriptors (taken from Ożańska-Ponikwia & Dewaele, 2012, p. 118)

HIGH SCORERS	GLOBAL DOMAINS	LOW SCORER
Curious, broad interests, creative, original, imaginative, untraditional	openness (O) Assesses proactive seeking and appreciation of experience for its own sake, toleration for and exploration of the unfamiliar.	Conventional, down-to-earth, narrow interests, inartistic, unanalytical
Organised, reliable, hard-working, self-disciplined, punctual, scrupulous, neat, ambitious, persevering	conscientiousness (C) Assesses the individual's degree of organisation, persistence, and motivation in goal-directed behaviour.	Aimless, unreliable, lazy, careless, lax, negligent, weak-willed, hedonistic
Sociable, active, talkative, person-oriented, optimistic, fun-loving, affectionate	extraversion (E) Assesses quantity and intensity of interpersonal interaction; activity level; need for stimulation; and capacity for joy.	Reserved, sober, aloof, unexuberant, task-oriented, retiring, quiet
Soft-hearted, good-natured, trusting, helpful, forgiving, gullible, straightforward	agreeableness (A) Assesses the quality of one's interpersonal orientation along a continuum from compassion to antagonism in thoughts, feelings, and actions.	Cynical, rude, suspicious, uncooperative, vengeful, ruthless, irritable, manipulative
Worrying, nervous, emotional, insecure, inadequate, hypochondriacal.	neuroticism (N) Assesses adjustment versus emotional instability. Identifies individuals prone to psychological distress, unrealistic ideas.	Calm, relaxed, unemotional, hardy, secure, self-satisfied

Openness describes a desire to search for and appreciate new experiences. DeYoung, Quilty and Peterson (2007) divided this trait between openness and intellect, with the former being closely related to its original meaning, the latter reflects a desire to gain intellectual stimulation.

Conscientiousness refers to one's ability to maintain motivation, organisation and a pro-active behaviour towards goal completion. High scorers tend to be hard-working, responsible and self-disciplined. DeYoung et al. (2007) split this trait between industriousness and orderliness.

Extraversion describes a desire to seek out social stimulation and engage in activities with others. Such individuals tend to be sociable and talkative. DeYoung et al. (2007) divided this trait between enthusiasm (positive emotion and sociability) and assertiveness (dominance and drive).

Agreeableness describes the extent to which one's actions reflect compassionate or antagonist behaviours. Agreeable individuals are friendly, trusting and kind-hearted, who care about the well-being of others (John & Srivastava, 1999). Agreeableness has been split across the dimensions of compassion and politeness (DeYoung et al., 2007).

Neuroticism is described as the extent to which individuals experience the world as distressing or threatening. Neurotic individuals tend to be highly strung, anxious and insecure. DeYoung et al. (2007) split the trait between volatility and withdrawal. While the former refers to the ability to control impulse and anger, the latter refers to one's vulnerability to anxiety.

Within such a taxonomy, these broad traits are comprised of narrow traits, also termed facets. Of particular interest to this project are the narrow traits of anxiety, curiosity and resilience. Anxiety can be defined as the frequency to which an individual is pre-dispositioned to feel nervous, worried or scared within their environment. Curiosity can be defined as a desire by an individual to show interest in new things, possess an open and receptive attitude and devote more attention to activities of interests. Resilience can be defined as an individual's ability to bounce back or recover from a stressful episode.

2.3 From Trait to State: A Change in Focus

Scholars have debated whether an individual's behaviour is predicted by their enduring individual characteristics or their momentary situational characteristics (Specht, 2017). The person-situation debate has served as the backdrop for personality research for the past half-century.

On the one hand, the trait theorists saw future behaviour as being predicted by long-standing thoughts, feelings and behaviours. Because it has been viewed that people behave similarly to themselves and differently from others, it was believed that underlying traits must be the reasons for this (Fleeson & Nofhle, 2008). Here, personality is conceptualised via a *nomothetic* approach, the aim of which is to “make general predictions about the population by examining between-person variation” (Beltz, Wright, Sprague, & Molenaar, 2016, p.447).

On the other hand, social-cognitive theorists believed the notions of traits to be fundamentally incorrect. Individual behaviour can be considered highly inconsistent across time and situation, and as a result, traits cannot be considered as good predictors of future behaviour (Mischel, 1977). These fluctuations in thoughts, feelings and behaviours *within* the same individual should be considered of great interest (Beckmann & Wood, 2020) and can be described and explained using personality states within an *idiographic* approach.

This debate is, however, all but now over, and a push towards resolution has gathered pace over the past decade (Fleeson & Nofhle, 2008). A compromise is an interactionist approach, allowing researchers to move towards a more complete understanding of *what* people do and *why* they do it (Funder, 2008). Studying the dynamic properties of personality allows for exploration of the underlying mechanisms accounting for changes seen in momentary cognitive, affective and behavioural states (i.e., thoughts, feelings and behaviours) together with those seen over a longer period (i.e., traits) and explore variability seen at the within-individual level over multiple timepoints (Beckmann & Wood, 2020). The rejection of a competitive view between ‘person’ and ‘situation’ has proven a vitally important step in furthering our understanding of personality.

The focus on states has allowed for a deeper understanding to be built regarding the value of within-personality variability in the dynamic properties of personality. Whereas within-person variability was once thought of as purely random error, and to be ignored, it is today thought of as being meaningful and systematic in some way (Beckmann, Wood, & Minbashian, 2010; Fleeson, 2001; Judge, Simon, Hurst, & Kelley, 2014). Indeed, this within-person variability can be studied at both the idiographic (i.e., what makes an individual unique) and at the group level (i.e., what do a group of individuals have in common) (Beckmann & Wood, 2020).

2.4 The Power of the Situation in State Personality

Historically, situations have always been understood as contributing to personality variability (e.g., Allport, 1961; Backteman & Magnusson, 1981; Mischel & Shoda, 1995). Nevertheless, theoretical and methodological limitations have until recently continued to limit the extent to which situations could be accurately conceptualised, taxonomised and measured (Hogan, 2009). One of the first conceptualisations of the relationship between situation and personality was made by Mischel and Shoda (1995) who theorised the Cognitive Affective Processing System Framework (henceforth CAPS). Mischel and Shoda (1995) posited that observable behaviour is triggered by the interplay of situational changes and our personality system (Beckmann & Wood, 2020). As everyone's personality is different, it can be assumed that individuals will behave differently to the same situation (Mischel & Shoda, 1998). Mischel and Shoda termed this process as *if... then* patterns. For example, when a situation is unfamiliar, anxiety is likely increased, which may then trigger shyness or neurotic trait-like behaviour. These patterns have been repeatedly shown to be stable across time in the same individual when faced with the same situation and serve as the cornerstone of within-person variability across different situations. Similarly, Fleeson (2007) argued that within-person variability in state personality can be systematically explained by changes in the perceived situation and that scholars should not simply ignore this variance as random error, but instead consider it psychologically meaningful. Fleeson termed this systematicity as a "contingency", defined as a systematic relationship between a given state and a given situation characteristic" (Fleeson, 2007, p. 830). The existence of such contingencies have been found in contexts including work (e.g., Beckmann et al., 2010; Huang & Ryan, 2011; Minbashian, Beckmann, & Wood, 2018) and the language learning classroom (Zhang, Beckmann, N., & Beckmann, J., 2019) and has provided an exciting area of research in the field of personality. Other literature (e.g., Sherman, Rauthmann, Brown, Serfass, & Jones, 2015) have also demonstrated how these situational contingencies can be used to account for personality characteristics, supporting that of Fleeson (2007).

2.4.1 Defining the situation

The above discussion has highlighted the importance of situations in accounting for within-person variance in trait-related personality states. However, defining what exactly a situation is, has traditionally, without a taxonomy or framework, proven troublesome for scholars and may serve as one reason why many have opted to avoid the topic (Hogan, 2009; Asendorpf, 2009, Rauthmann, Sherman, & Funder, 2015). Rauthmann et al. (2015) have argued that situations should be defined, taxonomised and measured using three kinds of situational information.

2.4.2 Cues, Characteristics and Classes

Situational cues describe the objective elements which make up a situation. Nettle and Gust (2015) use the acronym PEARLS to describe these cues; *p*erson (other people around someone), *e*vents (anything happening around someone), *a*ctivities (what people are doing), *r*oles (the formal and social roles that people inhabit), *l*ocation (space and time in which the situation is couched) and *s*tates (people's ambient thoughts, feelings, and desires). These cues do not possess any psychological meaning, and their objective nature means that the number and type of stimuli present would, in theory, be the same across all individuals. Questions regarding cues can be answered using 'wh' questions, for example, 'who', 'what' and 'where'. The study of cues has often been criticised (e.g., Hogan, 2009; Rauthmann & Sherman, 2016) due to the impracticality of capturing all cues in a given situation and the threat that some cues may be consciously observed by some individuals and not by others.

Situational characteristics differ from cues in that they do require psychological processing for they describe psychologically meaningful information (Ziegler, Horstmann, & Ziegler, 2019). These characteristics reflect the way humans process situational information and with it describe the situation. These characteristics can be interpreted as having some underlying mechanism which causes a direct interaction between the situation and person or has some discernible influence on an individual's behaviour (Edwards & Templeton, 2005). Because it requires psychological processing to be described, descriptions of situations are subjective and likely vary from individual to individual. For example, some individuals may find a project meeting to be stressful, while others find it productive (Parrigon, Woo, & Tay, 2016).

Situational classes aggregate situational cues and characteristics. Classification can be assigned based on similar cues (e.g., all situation is a classroom) or by similar levels of situational characteristics (Rauthmann et al., 2015). Ziegler et al. (2019) demonstrate this by classing a situation which is both intellectually challenging and cognitively demanding as 'learning'. They further note that there is currently no comprehensive taxonomy of situation class or context. The purpose of classes is to offer easy communication of situation type when information available may appear unclear or ambiguous (e.g., all situations termed as 'work situations').

2.4.3 Providing a situational framework

A primary goal of personality scholars has been to identify and organise situational characteristics into a systematic and generalisable format (Parrigon et al., 2016). Providing validated situational

taxonomies have not been as forthcoming as those developed in the *trait approach* (e.g., the Five-Factor Model). As Horstmann, Rauthmann and Sherman (2017) note that while more than 26 taxonomies have been put forward, many have lacked any form of validated measure. Consequently, scholars (e.g., Fleeson, 2007) were forced to use self-constructed and unvalidated measures to capture situational change (Rauthmann et al., 2014).

To address this, Rauthmann et al. (2014) proposed the ‘Situational Eight DIAMONDS’ taxonomy which divides situational characteristics into eight categories, which are as follows: **D**uty (Does work have to be completed?), **I**nтелlect (Does the situation require deep thinking?), **A**dversity (Is someone being blamed?), **M**ating (Are there potential romantic encounters?), **p**Ositivity (Is it a pleasant situation?), **N**egativity (Does the situation cause negative feelings?), **D**eception (Is someone being deceived?), and **S**ociality (Are there opportunities for social interaction?) (Rauthmann & Sherman, 2016). The taxonomy holds several advantages over its predecessors including its compatibility with other personality taxonomies, its ability to capture individual differences in situation perception and providing a description of momentary situations (Rauthmann et al., 2014). Since its publication, there has been a further four validated measures, which are as follows: The Situational Interdependence Scales (Gerpott et al., 2015); Situational Affordances as Adaptive Problems (Brown, Neel, & Sherman., 2015); CAPTION (Parrigon et al., 2016) and the Situation 5 (Horstmann & Ziegler, 2016).

2.5 Experience Sampling Methodology

The application of Experience Sampling Methodology (henceforth ESM) within a dynamic personality framework is today being realised. ESM offers a toolkit to follow participants intensely for short periods, capturing information on how they think, feel and behave *in the moment* (Larson & Csikszentmihalyi, 1983). This approach was posited nearly a century ago, with Allport (1937, p. 20) calling for a “novel and somewhat daring” methodology to capture not just between-person change (nomothetic), but within-person (idiographic) change also. Due to technological advancements, Allport’s vision is today a reality with an ever-increasing number of studies taking advantage of ESM to capture real-time data such as in the field of health, psychology, and work performance (Choe, Lee, N., Lee, B., Pratt, & Kientz, 2014).

The basic premise of ESM is that a participant is notified at a set or random time interval over a few days or weeks that they are required to complete a questionnaire. The questionnaire is usually sent during waking hours and should take less than 10 minutes to complete (Hektner,

Schmidt and Csikszentmihalyi, 2007). Questionnaires can refer to either the present time (i.e., how anxious do you currently feel?) or takes a retrospective perspective (i.e., how anxious have you been since the last alarm?). Notifications can be sent at set times (interval-contingent) or randomly (signal-contingent) or be based on a particular event (event-contingent). Signal-contingent is considered *true* ESM, with Conner and Lehmann (2012) noting that when investigating subjective experiences such as mood, signal-contingent timings are most suitable.

While early ESM studies were conducted with pen-and-paper, the advent of modern technology has allowed ESM as a tool to collect data prosper. Firstly, ESM has lower initial costs as software can be downloaded directly to the participants' smartphone. Secondly, ESM removes geographical boundaries, allowing findings to become more generalisable as researchers are no longer restricted by geographical location. Lastly, data collection is made effortless, and instead of needing to go to a laboratory, the researcher can download data directly from a server online.

Nonetheless, limitations in ESM with modern technology do exist. Firstly, ESM depends on battery-run devices and internet coverage means when participants devices have run out of battery, or they find themselves in an internet blackspot, data collection cannot occur resulting in a greater chance of missing data. Secondly, ESM tends to be data-collection heavy with continued notifications, sometimes proving annoying and inconvenient for participants. Lastly, some may see ESM studies as an invasion of privacy, particularly if one is using sensor-based data collection or where GPS positioning is required (Harari et al., 2017).

As a design, ESM studies may experience a higher attrition rate (Fisher & To, 2012), with participants becoming less compliant as time passes. This is likely due to the repeated measurement points required in an ESM study and the intrusive nature of notifications as noted above. Researchers can, however, limit this by adjusting designs accordingly, for example enabling notifications to be sent between 9 am and 5 pm or keeping the data collection window open for a little while after the first notification.

2.6 Principles of Personality Development

Much attention has been given to understanding the mechanisms which cause long-term broad trait change. The development of the Neo-Socioanalytic Model (see Table 2) looked to embody “the structure and content of personality as well as a set of principles of personality development and mechanisms of continuity and change” (Roberts & Nickel, 2017, p. 161). In light of previous

literature (e.g., Zimmermann & Neyer, 2013), of relevance to the thesis is the Maturity Principle, The Social Investment Principle and The Correspondive Principle.

Regarding the Maturity Principle, the basic premise pertains that with age, individuals tend to become more agreeable, conscientious and less neurotic. Literature (e.g., Donnellan, Conger, & Burzette, 2007; Donnellan & Lucas, 2008; Josefsson et al., 2013) has consistently found such patterning regardless of culture or language indicating strong support for the Maturity principle. The Social Investment Principle posits that personality changes because there are changes in an individual's social role which require subsequent behavioural changes (e.g., being a parent or an employee), with longitudinal data (e.g., Bleidorn et al., 2013; Lehnart, Neyer, & Eccles, 2010; Leikas & Salmela-Aro, 2015) tending to support this principle. This principle has been used to, in part, explain why individuals mature with age (Roberts & Nickel, 2017).

Concerning the Correspondive Principle, this views personality development of a particular trait as a direct result of experiencing certain environments, directly related to that specific trait. For example, individuals may take up a leadership position because they are dominant as individuals and will become more dominant through their experiences as a leader (Caspi, Roberts & Shiner, 2005). The strength of evidence is lower than the aforementioned principles, but nonetheless, evidence of Correspondive Principle has been found within the work environment (e.g., Li, Fay, Frese, Harms, & Gao, 2014), educational settings (e.g., Roberts & Robins, 2004) and daily activities (Allen, Magee, Vella, & Laborde, 2017).

Table 2: Principles of Personality Development (Roberts & Nickel (2017, p. 162)

Principle	Definition	Strength of evidence
Cumulative Continuity Principle	Personality traits increase in rank-order consistency until midlife	Strong
Maturity Principle	People become more socially dominant, agreeable, conscientious, and emotionally stable with age	Strong
Social Investment Principle	Investing in social institutions, such as age-graded social roles, outside of the self is one of the driving mechanisms of personality development in general and greater maturity in particular	Good
Corresponsive Principle	The effect of life experience on personality development is to deepen the characteristics that lead people to those experiences in the first place	Moderate
Plasticity Principle	Personality traits are open systems that can be influenced by the environment at any age	Strong
Role Continuity Principle	Consistent roles rather than consistent environments are the cause of continuity in personality over time	Weak
Identity Development Principle	With age, the process of developing, committing to, and maintaining an identity leads to greater personality consistency	Weak
Niche-Picking Principle	Through their personality traits, people create social environments and paths in their lives that help maintain their current trait levels	Weak

2.7 Personality Development Between 18 – 25: The Age of Wonder

Identity development has been consistently found to be a catalyst for personality change (Klimstra & van Doeselaar, 2017) as individuals go through a state of identity-confusion and exploration. As Arnett (2000) notes, for many in western cultures, this occurs between the ages of 18 and 25, a period he calls emerging adulthood, falling between adolescence and fully-fledged adulthood. For many, this period coincides with life events such as entering university, completing a sojourn or entering the workforce.

For those in emerging adulthood, the non-committal to adult-graded roles at 18 is likely driven by the decision to study in tertiary education. Such non-committal is demonstrated by delayed changes in personality traits associated with maturity, as demonstrated by both cross-sectional and longitudinal research (e.g., Bleidorn et al., 2013, Lüdtke, Roberts, Trautwein, & Nagy, 2011; Roberts et al., 2006a). Golle et al. (2019) found those who choose a vocational path, as opposed to an academic path, matured at a faster rate and showed a decreasing interest in domains such as social and enterprising interests. Sojourning, on the other hand, maybe considered to foster maturational processes for learners must display behaviours associated with adaptiveness in order to successfully integrate, reflected by changes in agreeableness and neuroticism (Zimmermann & Neyer, 2013).

As emerging adulthood is seen as a time of social exploration, changes in trait openness are also to be expected. Soto, John, Gosling and Potter (2011), for example, found significant mean score increases in openness during ages 18-22. This change was likely driven by entry into tertiary education, whereby individuals are met with intellectually stimulating challenges and socialising with new friendship groups. This finding was also supported by Lüdtke et al. (2011) who found that regardless of career path (studying, vocational training, or work) all participants between the ages of 19-22 experienced significant mean level changes in openness. Zimmermann and Neyer (2013) also found sojourners' openness levels to increase while abroad mediated by the changes in everyday social behaviour and the forming of new relationships while abroad.

2.8 Personality and Life Events

It is today agreed upon that genetics, natural maturation processes and the environment all influence the stability and malleability seen in personality traits. Within the *environmental* component is the occurrence of life events. Such events have been defined in differing ways and can refer to both time-discrete events (e.g., birth, marriage) or slow and continuous (e.g., transition into adulthood) (Luhmann, Orth, Specht, Kandler, & Lucas, 2014).

Embedded within the notion of life events are the terms *self-selection* and *socialization*, both of which are relevant to the broader objectives of Study 2 (see Chapter 1). Self-selection effects refer to the notion that an individual's trait make-up serves as a mechanism for encountering a certain situation or event. Such differences in trait make-up may account for the observed "systematic differences in the occurrence of these events" (Zimmermann & Neyer, 2013). From a sojourning perspective, self-selection effects would posit that certain individuals are more likely to choose to study abroad when the opportunity is voluntary (see section 2.9.1.2). Socialization effects, on the other hand, can be viewed as the effect the environment has on an individual's personality, with these experiences being viewed as life events.

Specific to Study 2, is the life event of educational transitioning, whether home or abroad. Sojourning, for example, involves a complete upheaval of the cultural norms known and requires psychological adaptation in order to find belonging. Brown and Holloway (2008, p. 33) see this transition as "one of the most traumatic events in a person's life". Research has repeatedly viewed the sojourning experience as transformative, one which can have great academic, personal and social consequences (Andrews, Page, & Neilson, 1993; Leong & Ward, 2000; Searle & Ward, 1990). Living abroad can offer creative freedom together with the ability to meet new

people and try new experiences. Research (e.g., Niehoff et al., 2017; Zimmermann & Neyer, 2013) has found sojourning to be a valid life event, one which can result in personality change. At-home study, particularly that of final year undergraduate study, has been relatively unexplored in relation to being a valid life event. Instead, research has tended to focus on the transition from high school to university (e.g., Bleidorn, 2012; Lüdtke et al., 2011).

2.9 The Role of Learning Context in Accounting for Personality

Development

This thesis explores the role learning context plays in accounting for the observed changes seen in a range of measured outcomes. In this section, a particular focus is given to the outcome of personality, with the following literature review providing relevant evidence regarding the value of each of the learning contexts in relation to measurable personality change.

2.9.1 Study abroad

Historically, personality research within a sojourning context has received little scholarly attention. Despite this, personality has long been thought of as predictive of successful integration and acculturation abroad (Bakalis & Joiner, 2004; Gu & Maley, 2008; Leong, 2007; Williams, 2005), while personal growth has been considered a primary goal of such experiences. (Anderson, Lawton, Rexeisen, & Hubbard, 2006; Mather, Karbley, & Yamamoto, 2012). The following narrative review breaks down past research, exploring how personality can both influence programme entry and develop as a result of participation.

2.9.1.1 Self-selection effects

In countries and disciplines where participation onto an exchange programme is voluntary; self-selection effects may be present as explored in Section 2.8. To date, the available literature (e.g., Bakalis & Joiner, 2004; Greischel, Noack, & Neyer, 2016; Gu & Maley, 2008; Niehoff et al., 2017; Zimmermann & Neyer, 2013) has indeed demonstrated that those who choose to study abroad, are systematically different to those who choose to remain at-home. Such literature has tended to find that those who choose to study abroad are more open, conscientious, extraverted, agreeable and less neurotic than those who remain at-home, suggesting that an individual's decision to undertake a sojourn experience is likely influenced by their personality make-up. Self-selection effects within the prism of study abroad research (e.g., Zimmermann & Neyer, 2013) may be considered of less importance in this thesis, given that for many, the experience is compulsory and as such, cannot self-select themselves to partake in a study abroad experience.

Nonetheless, it should be remembered that these same individuals have self-selected themselves to study a degree where the experience is offered and as such, it cannot be ruled out that the study abroad was a determining factor in their decision to undertake the degree in the first place.

2.9.1.2 *Socialization effects (broad personality traits)*

Section 2.9.1.2 looks to explore the literature concerning observed socialization effects (see section 2.8) within the sojourning context, which to date consists of very few studies.

Of the most robust, due to study design, is that of Zimmermann and Neyer (2013) and Niehoff et al. (2017), for both implemented a quasi-experimental design, comparing an intervention group (i.e., abroad) with an undergraduate comparison group (i.e., at-home) over one semester and one year, respectively. Both these studies consisted of a German L1-speaking undergraduate sample, for whom the study abroad experience was voluntary. In each of the aforementioned studies, broad trait personality (e.g., openness, conscientiousness etc.) was measured using the German translated version of the 'Big Five Inventory', and while broad trait personality was captured in up to four waves of data collection in Zimmermann and Neyer (2013), participants in Niehoff et al. (2017) were measured twice (i.e., pre/post-test).

A consistent finding across both studies was that agreeableness, in particular, benefited from a period abroad. In Zimmermann and Neyer (2013) both short-term sojourners ($p = .18$; $d = .10$) and long-term sojourners ($p = .04$; $d = .13$) became significantly more agreeable than their at-home peers in the first semester, while similarly Niehoff et al. (2017), found sojourners to experience significantly greater change in agreeableness ($p = <.01$) than their at-home peers over one semester. While in the traditional sense, the observed effect sizes in Zimmermann and Neyer (2013) could be considered small, given the slow-natured change of broad trait personality, the observed change should be considered meaningful.

While Zimmermann and Neyer (2013) also observed significant socialisation effects in both openness ($_{short} d = .23$; $_{long} d = .12$) and neuroticism ($_{short} d = -.13$; $_{long} d = -.16$), no such effect was observed for these traits in Niehoff et al. (2017). Neither study demonstrated sojourning to foster the development of extraversion or conscientiousness.

Gresichel, Noack, & Neyer (2016), investigated personality development in German high schoolers who either completed a year abroad ($n = 457$) or remained at-home ($n = 284$). All

participants were assessed three times across one year. The findings demonstrated that across the academic year, adolescent sojourners showed enhanced increases in openness, agreeableness, and emotional stability. No sojourning effect was found for conscientiousness and extraversion supporting the findings of Zimmermann and Neyer (2013) and Niehoff et al. (2017).

As aforementioned, due to the inclusion of a stay-at-home comparison group, the above studies provide the most robust causal evidence that the sojourn experience accounted for the observed personality change. However, finding a suitable comparison group is not always possible or feasible, and in such cases, studies must revert to a one-group pre-post-test design (e.g., Arvidsson, Eyckmans, Rosiers, & Lundell, 2018; Schartner, 2016; Tracy-Ventura, Dewaele, Köylü, & McManus; 2016). Such studies have tended to find change in emotional stability in particular. Tracy-Ventura et al. (2016) for example, observed personality change in 58 British undergraduate students partaking in a year abroad. Students were shown to return home significantly higher in emotional stability ($p = .042$; $d = -.30$) with all other traits remaining relatively unchanged. However, given the study design implemented, it cannot be determined as to whether the sojourning experience is causally attributed to this change. That is to say, it cannot be ascertained as to whether the same participants would have demonstrated similar changes at-home and it is not possible to differentiate this treatment effect from another effect. In all of the aforementioned studies, the evidence provided has explored average change in participants over time (i.e., nomothetic). The field of personality psychology, has, however, become interested in understanding the role of individual differences in personality change trajectories (e.g., Specht, 2017). While no study, to the best of my knowledge, has explicitly examined individual differences in personality change over time while abroad, Stronkhorst (2005) did report substantial individual differences in the extent of development of multicultural traits. Stronkhorst (2005) found between 35% and 45% of the participants showed improvement in the areas of Cultural Empathy, Open-mindedness and Flexibility, with others showing little change or regression. As such, it was concluded that the study abroad experience was heterogeneous, with individuals finding the study abroad experience to be more valuable than others, in relation to personality development. Understanding the extent of these differences is, therefore, an avenue of further research.

2.9.1.3 Socialization effects (narrow personality traits)

Turning attention to the narrow personality traits of anxiety, curiosity and resilience, the lack of comparative studies poses issues in determining just how facilitative a sojourn experience is in

developing these traits. Instead, data captured on these traits have typically been conducted retrospectively, upon completion of the sojourn experience, through means of focus groups and interviews. Such reliance on these approaches limits the validity of the evidence discussed for findings may be subject to recall bias and the perceived impact of the sojourn experience on these traits may be subjectively influenced by an individual's satisfaction towards the experience. Consequently, the strength of evidence must be tempered by such limitations.

2.9.1.3.1 *Anxiety*

Concerning *anxiety*, sojourners perceive anxiety to be highest at pre-departure and then decrease over time. Savicki and Price (2017) for example, analysed blogs of 36 sojourning students, finding that words associated with anxiety (e.g., afraid, alone, pity) were most ascribed before departure coinciding with the uncertainty ahead and then decreasing thereafter. Studies comprising of retrospective focus groups (e.g., Forsey, Broomhall, & Davis, 2012; McLeod & Wainwright, 2009; Tracy-Ventura et al., 2016) have observed sojourners to perceive a reduction in anxiety across the study abroad period as they become better able to deal with unfamiliar or difficult situations. For example, in Tracy-Ventura et al. (2016), students repeatedly remarked how they found overcoming difficult situations empowering, promoting self-confidence and self-efficacy.

2.9.1.3.2 *Curiosity*

Regarding *curiosity*, manifestations of this trait have often been viewed through the lens of intercultural awareness and intercultural competence (Deardorff, 2006) as opposed to being studied as a defined psychological construct (Chieffo & Griffiths, 2004; Ingraham & Peterson, 2004; Soria & Troisi, 2014). There is, to the best of my knowledge, no study which measures curiosity development quantitatively during a sojourn. Nonetheless, findings have consistently shown that sojourning promotes intercultural awareness. In a meta-analysis conducted by Haas (2018), sojourning was found to have a significant impact on an individuals' cultural competence ($d = 0.56$), reflecting an increased desire and understanding of other cultures. Tracy-Ventura et al. (2016) noted how some individuals returned home with a sense of eagerness to explore other countries and interact with individuals from other countries. Moreover, Forsey et al. (2012) found sojourners to return home with expanded horizons. Students now felt 'intellectually' linked with the world and felt more 'global' as result of sojourning. Around half ($n = 7$) of the interviewees felt they came home more open-minded after experiencing another way of life.

2.9.1.3.3 Resilience

Lastly, regarding *resilience*, little is known regarding the extent to which sojourning improves a learner's resilience, which is perhaps surprising given its close relationship with learners' coping mechanisms, ability to adapt and homesickness (Suanet & Van de Vijver, 2009; Thomas, 2018). Earnest, Rosenbusch, Wallace-Williams and Keim (2016) undertook a quasi-experimental design, similar to that of Zimmermann and Neyer (2013) comprising of sojourners ($n = 22$) and non-sojourners ($n = 82$). A significant interaction effect ($p = <.001$, $\eta^2 = .26$) was observed for emotional resilience, representing a large-sized effect, and indicating *sojourners* to become significantly more resilient than *non-sojourners*. With such limited evidence, however, it is difficult to come to any firm conclusions regarding the impact of sojourning on resilience development.

2.9.2 Undergraduate study

In contrast to the sojourning context, section 2.9.3 looks to explore studies which have discussed personality development in relation to the at-home context. This focus is important, given that the purpose of the thesis as a whole is to compare change in the sojourning context with that of at-home undergraduate study, with a particular focus on the 3rd year of study. To date, however, little scholarly researched has been conducted on capturing personality change between first and final years of university study and even less so has focused specifically on changes in the final year in particular, which is the purpose of this thesis (Study 2).

2.9.2.1 Socialization effects (broad personality traits)

To the best of my knowledge, only Lüdtke et al. (2011) has provided robust evidence regarding the extent to which a university setting can foster personality change. Comparing individuals who were transitioning from high school to university ($n = 1,179$) with those transitioning to work or vocational training ($n = 597$), two significant interaction effects were found. Firstly, those who attended university were found to become significantly more agreeable over time compared to those who worked, while those who worked became significantly more conscientious over time compared to the university group.

More frequently, pre/post-test studies have been undertaken, which provide more descriptive evidence regarding the extent of change, as opposed to the cause of such change. For example, Vollarth and Torgersen (2000) examined 119 students, who were measured at the beginning and end of their university programme. The results suggested that neuroticism decreased over time; agreeableness and conscientiousness significantly increased, while extraversion and openness

showed no change over time. Conversely, Schurer, Kassenboehmer and Leung (2015) found extraversion to increase in university students by one-third of a standard deviation, with this observation being explained by the assumption that university fosters extraverted tendencies as it encourages participation in extracurricular and social activities.

In sum, there is mixed evidence regarding the extent to which university can foster personality change and is perhaps a reflection of the variability in environments experienced by students worldwide (e.g., accommodation types). The lack of comparative studies is also limiting when looking to identify causation and the limited number of studies available also inhibit generalisability of findings, suggesting further research is warranted.

2.9.2.2 *Socialization effects (narrow personality traits)*

2.9.2.2.1 Anxiety

To date, studies focusing on the notion of anxiety within an undergraduate population have typically subsumed the construct within the term well-being, with such studies being explored in Section 2.10.2. Nonetheless, university study may be considered a particularly anxious and stressful period in light of the academic, social, and financial pressures which come with tertiary education (Husky, Kovess-Masfety, & Swendsen, 2020). Moreover, typically studies within an undergraduate sample investigating anxiety have tended to avoid viewing it as a holistic psychological construct instead narrowing it down to its lower-level facets such as Test Anxiety (e.g., Chapell et al., 2005; Roos et al., 2020); academic anxiety (e.g., Yang, Asbury, & Griffiths, 2019) and foreign language anxiety (e.g., Thompson & Lee, 2014). There is to date, little research which has captured the longitudinal nature of anxiety change in the final year of academic study.

2.9.2.2.2 Curiosity

Regarding *curiosity*, a university setting has been hypothesised to spark curiosity by introducing individuals to a diverse population while simultaneously providing a stimulating intellectual environment (Robins et al., 2001). While curiosity has repeatedly been shown to be a strong predictor of academic achievement at university (e.g., von Stumm, Hell, & Chamorro-Premuzic, 2011), to the best of knowledge, no study has explored curiosity development while at university.

2.9.2.2.3 Resilience

In relation to *resilience*, Bleasdale and Humphreys (2018) investigated resilience in a sample of 2nd-year undergraduates at Leeds University. Fifty-five students were interviewed once during their

academic study and asked to retrospectively reflect on resilience changes during university. Many indicated their resilience to increase during the two years spent at university. Individuals had found settled friendship groups and had now got into a routine regarding their studies. While informative, interviews are limited by bias, minimising the validity of the responses.

In summary, little longitudinal research has been conducted on the narrow traits and questions remain regarding both *how* and *why* personality fluctuates over time. Nonetheless, it should be noted that from a sojourning perspective, while quantitative approaches have been limited in their use, there is much research which has discussed personality using anthropological and ethnographic methodologies and it is the purpose of section 2.10 to discuss these in more detail.

2.10 Personality, Study Abroad and its Differing Paradigms

Study 2 sees personality through the prism of personality science, using the framework of the 'Big Five' (i.e., Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism). Within the field of study abroad, personality has also been theorised and examined within "ethnographic and poststructuralist thinking", using predominantly qualitative measures (Mitchell, Tracy-Ventura, & McManus, 2015, p. 12).

Driven by the need among cross-cultural psychologists for a personality measurement which was more sensitive to contextual variables, Van der Zee and Van Oudenhoven (2000) developed the Multicultural Personality Questionnaire (MPQ). While its dimensions can be mapped on to the 'Big Five', they are more specific towards multicultural effectiveness, and in doing so were considered "specifically tailored to grasp individual differences predictive of multicultural attitudes and multicultural success" (Leone, Van der Zee, Van Oudenhoven, Perugini, & Ercolani, 2005, p. 1450). The dimensions measured are as follows: Openness (strongly correlated with openness and extraversion); Emotional Stability (strongly correlated with neuroticism and agreeableness); Social Initiative (strongly correlated with agreeableness and conscientiousness) and lastly, Flexibility (strongly correlated with openness and extraversion). The tool has been utilised in several studies (Leong, 2007; Tracy-Ventura et al., 2016) and places the construct of personality within the prism of cross-cultural and socio-psychological adaptation. When utilising the MPQ, findings have tended to show sojourners to become more emotionally stable and open over time while abroad (Leong, 2007; Tracy-Ventura et al., 2016) and are line with findings presented in section 2.9.1.

Similarly, the narrow traits of anxiety, curiosity and resilience have all been examined within the interculturality literature.

Anxiety has been explored within the remit of intercultural communication apprehension (ICA) defined by Neuliep and McCroskey (1997, p. 148) as “the fear or anxiety associated with either real or anticipated communication with people from different groups, especially cultural and/or ethnic groups.” High levels of ICA are likely to inhibit one’s desire to engage in intercultural interactions, minimising the extent to which a sojourner can successfully integrate and adapt to the new community (Neuliep, 2012).

Curiosity serves as a cornerstone of measuring intercultural competencies (Houghton, 2014) and has been recognised as a component of Byram’s (1997) model of Intercultural Communicative Competence (ICC). Curiosity may thrive during a study abroad experience, for individuals have the potential to be offered new experiences and interact with new members of the host community. Research indicates a link between curious behaviours (e.g., inquisitiveness; tolerance of ambiguity and suspension of judgement) and intercultural effectiveness (Deardoff, 2009).

Resilience, particularly that of emotional resilience, has been linked with cross-cultural adaptability (Taguchi, 2015). Emotional resilience assesses an individual’s attitude towards coping with stress, is closely aligned with the psychological definition (i.e., ability to ‘bounce back’ from stress) and has been continuously shown to be highly correlated with greater adaptability (e.g., Fong, 2020; Shaftel, J., Shaftel, T., & Ahluwalia, 2007). Those high in resilience are better able to deal with the psychological challenges presented by living in a foreign country and cope with unfamiliar experiences. Resilience has also been explored alongside the construct of agency (Jackson, 2008; Gu, Schweisfurth, & Day, 2010). In summary of such research, it may be said that those who exercise their own agency and demonstrate resilience have more successful adaptive strategies and find it easier to successfully integrate into the host community.

As earlier noted, the use of ethnography when exploring personality change has been a common practice, borne out of the field of anthropology. This methodology allows scholars to examine cultural aspects of a host community from an inside perspective and maybe utilised through direct observation, blogs and keeping journals. To date, the LARA project (Roberts, Byram, Barro, Jordan, & Street, 2001) and the Special English Stream (Jackson, 2010) serve as two large-scale ethnographic projects regarding the study abroad experience. Within the study of

intercultural competence, ethnographic research has the advantage of moving away from focusing on factual knowledge gain and allowing insights into the process of how intercultural competence develops (Bateman, 2002). Through providing an opportunity to interact with host members of the community, students gain an insider's perspective of culture and learn themselves by reflecting on their own beliefs, values and customs (Lee, 2012). Ethnography can also present insights into how culture and new opportunities are experienced 'in the moment' and in doing so provide distinctive contributions to our understanding of how a period abroad is lived by sojourners. Ethnographic research has tended to find study abroad to facilitate language learning and personal growth (e.g., Bacon, 2002; Kinginger, 2011).

To summarise, section 2.10 has recognised that personality within the field of study abroad has been examined in several ways. While Study 2 explores personality within the prism of personality science, others have linked personality with interculturality, recognising it as an important component in the development of intercultural competencies. All such research, regardless of theoretical thinking, can provide important insights and provide a more comprehensive account of how sojourning can impact personality.

2.11 Well-being and the Role of Learning Context in Education

Understanding fluctuations in well-being are becoming increasingly important for tertiary institutions who have seen a sharp decrease in student well-being over the past few decades (e.g., Bewick, Gill, Mulhern, Barkham, & Hill, 2008; Bewick, Koutsopoulou, Miles, Slaa, & Barkham, 2010; Nevens & Hillman, 2019; Stewart-Brown et al., 2000). Well-being has been referred to by Ryan and Deci (2001) as optimal psychological functioning and as a construct, has typically been divided between subjective well-being and psychological well-being. While the former relates to the pursuit of happiness and a pleasant life, the latter focuses on the fulfilment of human potential and a meaningful life (Chen, Jing, Hayes, & Lee, 2013). At the general construct level, these two constructs are highly related; however, as Chen et al. (2013) note, the factor structure of each is highly individualised.

Personality traits have repeatedly been shown to share moderate to strong correlations with well-being (Lucas, 2018), and Vittersø and Nilsen (2002) have found that approximately 50% of the variance in well-being can be explained by personality traits. McCrae and Costa (1991) presented two frameworks as an explanation as to why personality and well-being are highly related. The first is the *temperament model* which predicts that well-being and personality traits are associated

because both are linked with consistent affective experiences. The second is the *instrumental model*, which views that differing daily behaviours are predicted and affected by traits, which in turn are associated with affective experiences (Howell, Ksendzova, Nestigen, Yerahian, & Iyer, 2016). Such robust associations have led some to argue (e.g., Schmutte and Ryff, 1997) that as well-being and personality share common underpinnings, the two constructs are tautological.

Theoretically, personality traits should also influence well-being indirectly. Lower-order traits such as warmth, friendliness and gregariousness should foster positive social relations and a strong support network, which in turn should positively impact well-being. Similarly, personality traits may disposition individuals to behave in a certain manner which consequently impacts their perceived well-being. Neurotic individuals, for example, maybe at a disposition to view their current experience as anxiety-inducing or stressful, inhibiting positive well-being (Røysamb, Nes, Czajkowski, & Vassend, 2018). To date, evidence has shown a well-established relationship between extraversion and positive affect (i.e., well-being); and neuroticism and negative affect (e.g., DeNeve & Cooper, 1998; Steel, Schmidt, & Schultz, 2008). Conscientiousness has shown more mixed results. While conscientious tendencies such as *orderliness*, *dutifulness* and *cautiousness* are likely to have a positive effect on well-being (Friedman, Kern, & Reynolds, 2010), those with extreme high conscientiousness may experience fewer positive outcomes (Carter, Guan, Maples, Williamson, Miller, 2016). After all, those high in conscientiousness are likely *perfectionists* and demonstrate an obsession with routine, order and cleanliness, displaying symptoms of obsessive-compulsive disorders (Samuel, Riddell, Lynam, Miller, & Widiger, 2012). Trait agreeableness and openness, on the other hand, appear to influence well-being to a lesser extent (e.g., DeNeve & Cooper, 1998; Steel et al., 2008).

It is of pertinence to the aims of objectives of the thesis is to briefly discuss research to date which has explored well-being in each of the relevant contexts. The discussion first begins with the sojourning context, followed by the at-home context.

2.11.1 Study abroad (sojourners)

While an increasing number of students undertake a period abroad, the extent to which the experience fosters positive well-being is yet to be explored in a European (Lees, 2020; Potter, 2020) and American (Poyrazli & Mitchell, 2020) population, and is perhaps indicative of the broader literature which has tended to focus on implementing interventions to reduce negative affect (e.g., anxiety, distress, Blackie, Roepke, Forgeard, Jayawickreme, & Fleeson, 2014).

During a period abroad, positive well-being is threatened by variables which Yakunina, Weigold and McCarty (2011, p. 68) term “acculturation stressors”. These stressors are diverse and varied (e.g., identify conflict, language barriers, weather) and impact negatively on one’s psychological adaptation to the new environment. In instances where this adaptation is hindered, situations become stressful and challenging. Any prolonged exposure to these situations will likely result in culture shock and consequently, negative well-being (Soto, 2015). These threats are likely highest in the first four to six months abroad where learners must overcome adaptation hassles which can lead to mood disturbance (Furukawa & Shibayama, 1993).

Conversely, research has shown that overcoming challenges experienced during the study abroad period can lead to positive well-being. As Furnham and Bochner (1986, p. 47) note, overcoming these challenges can “make a person more adaptable, flexible, and insightful”, which may buffer stress and anxiety when learners are met with a future challenging event. Typically, this view has been evidenced by retrospective data collection, whereby learners provide testimonies upon return home. For example, Potter (2020) surveyed 33 individuals at post-sojourn with many participants perceiving their well-being to increase after a) dealing with everyday tasks (e.g., opening a bank account), b) overcoming a negative event, and c) successfully integrating into the host community. Similarly, Milstein (2005) retrospectively questioned 212 sojourners who partook in a Japanese exchange programme. 95.5% of the sample noted a significant perceived improvement in self-efficacy, which, according to Natovová and Chylová (2014) is a marker of an individuals’ psychological well-being.

For those living in a non-L1-English speaking country, linguistic competencies may directly influence perceived well-being. From one perspective, limited proficiency may inhibit successful integration and adaptation, which will likely inhibit well-being (Churchill & Dufon, 2006). Savicki, Arrué and Binder (2013) tested this hypothesis by exploring two groups of learners. One group of learners studied in Oviedo and required a particular linguistic threshold to be accepted. The other studied at Vienna and included learners with no or limited second language (henceforth L2) German proficiency. In other measured respects, the two groups were the same. At the end of the period abroad, those who studied in Spain did not perceive significantly higher well-being than those in Vienna, leading to Savicki et al. (2013) to conclude that prior ability to use the L2 of the host community had little influence on perceived well-being.

Conversely, language learners may experience inflated well-being over time as they become more proficient in their L2. The findings of Potter (2020) provide tentative evidence that this

relationship may hold true. 82% of individuals noted that increased linguistic competence had a positive impact on their well-being. Potter (2020, p. 25) notes “Whether consciously or not, that ability to communicate in another language, and possibly because it is a non-native language, engenders a sense of well-being.”

2.11.2 At-home instruction (non-sojourners)

Psychological well-being changes have typically been investigated during the transition phase from home to university (e.g., Andrews & Wildig, 2004; Cooke, Bewick, Barkham, Bradley, & Audin, 2006), where findings indicated that well-being worsened as students began university, dropping substantially lower than that found pre-ceding entry. Moreover, research has generally investigated well-being using one-off cross-sectional surveys, many of which have not been administered at differing timepoints throughout the same cohort. However, Bewick et al. (2010) conducted the first longitudinal study regarding well-being change in UK undergraduates who began their studies between 2000 and 2002, respectively. Well-being was measured using the GP-CORE and administered twice in each undergraduate year, once in autumn and once in spring. One hundred eighty-seven students completed the questionnaire at each timepoint. The findings indicated that psychological well-being tended to decrease throughout their study. Average well-being was at year three was approximately twice as worse as measured at pre-registration (baseline) and 30% worse than that measured at T2 (year one). In year three, psychological well-being was found to be significantly lower in the second semester compared to the first, regardless of gender. This finding goes against other research (e.g., Cooke et al., 2006) which found women to have on the whole more negative well-being while at university than males.

In summary, the available literature tends to indicate that sojourning may benefit well-being through overcoming adversity and fostering positive self-esteem. Conversely, at-home study appears to inhibit positive well-being with cumulative evidence indicating tertiary study to be a stressful and anxiety-inducing period. Nonetheless, the described methodological limitations continue to compromise our full understanding of how well-being fluctuates over time both in the abroad and at-home learning contexts, and more longitudinal research is required to explore these, the purpose of which Study 2 looks to achieve.

This chapter has explored the constructs of personality and well-being. Chapter 3 explores Second Language Acquisition, which is explored in both Study 1 and Study 2 of this thesis.

Chapter 3: Second Language Acquisition

Second Language Acquisition (henceforth SLA) defines the methodical study of any learnt language other than your first (Macaro, Murphy, & Vanderplank, 2013). The field developed in the 1960s behind the backdrop of the ‘nature vs nurture’ debate, which looked to ascertain whether learning language was genetically driven (i.e., nature) or through interaction with the environment (i.e., nurture) (Gass & Selinker, 2008). The 1970s and 1980s saw increasing attention paid to teaching pedagogy and the notion of the ‘good language learner’ (Rubin, 1975). Recent research has taken a more multifaceted approach in research direction, incorporating constructs and methodologies from other disciplines such as Sociology and Psychology which continue to drive the field forward. SLA research interests itself with the processes of language acquisition, which can be broadly defined as learner-external (e.g., the learning environment) and learner-internal (e.g., cognitive ability, age and personality). It is, therefore, the purpose of this chapter to firstly, evaluate the role sojourning plays in facilitating language acquisition, and secondly, to explore factors associated with development rate, with a particular emphasis given to personality.

3.1 Study Abroad: The ‘Golden Ticket’ for Linguistic Change?

Public opinion has long viewed exchange programmes as providing the optimal environment for language acquisition (Barquin, 2012; Hessel, 2017). Such claims stem from the notion that living abroad offers the most immersive L2 experience, one which can offer both formal (in-class language instruction) and informal (out-of-class language use) learning opportunities (Llanes & Muñoz, 2013). These claims are further supported by SLA theory which purports the benefits of study abroad through affording learners’ optimal opportunity to listen (Input Hypothesis, Krashen, 1985), to speak (Output Hypothesis, Swain, 1985) and to interact actively and negotiate for meaning in the L2 (Interaction Hypothesis, Long, 1996).

However, whether these programmes are today as immersive as they once were, is debatable. As explored in Section 1.2, advances in communication, together with the rise of English as a Lingua Franca means for L1 English speakers at least, “abroad is less abroad than it once was” (Coleman & Chafer, 2010, p. 165). Understanding and researching the notion of an ‘immersion myth’ is important, given that for students and stakeholders alike, improving linguistic proficiency is considered the goal for many programmes (British Council, 2005).

Today, the field of study abroad within SLA can be considered to have reached maturity (Tullock & Ortega, 2017) due to the vast amount of literature published on the topic. Since Carroll's (1967) seminal study detailing the benefits of a year abroad on linguistic development, research in the area has boomed, reflecting the continuing internationalisation of education (Jackson, 2018). According to the literature, undertaking an overseas exchange appears to facilitate the skills of oral fluency (e.g., Freed, 1995; Freed, Segalowitz & Dewey, 2004; Juan Garau & Pérez-Vidal, 2007; Kim et al., 2015; Lindseth, 2010; Llanes & Muñoz, 2009; 2013; O'Brien et al., 2006; Mitchell et al., 2017a; Mora & Valls-Ferrer, 2012; Segalowitz & Freed, 2004; Serrano et al., 2012); listening (e.g., Allen, 2002; Beattie, Valls-Ferrer, & Pérez-Vidal, 2014; Kinginger, 2008; Llanes & Muñoz, 2009), pragmatic competence (e.g., Félix-Brasdefer & Hasler-Barker, 2015; Shively, 2011, 2013; Warga & Schölmberger, 2007), and vocabulary size (e.g., Dewey, 2008; Ife, Vives Boix, & Meara, 2000; Jiménez Jiménez, 2010; Milton & Meara, 1995). Writing accuracy and fluency (e.g., Godfrey, Treacy, & Tarone, 2014; Pérez-Vidal & Juan-Garau, 2009; Mitchell et al., 2017a; Serrano et al., 2012; Wu & Zhang, 2017) and reading skills (Gomes da Costa, Smith, & Whitely, 1975; Kinginger, 2008) have demonstrated more mixed results, while syntactic competencies (Howard, 2006, 2008; Rees and Klapper, 2007) and pronunciation (Avello and Lara 2014, Díaz-Campos 2004) have also proven inconclusive in their conclusions. In sum of the available literature, linguistic gains tend to be subtle, highly individualised and highly compartmentalised across linguistic domains (Hessel & Vanderplank, 2018), and as a result, conclusions drawn regarding language acquisition in a sojourning context tend to be tentative (e.g., Tullock & Ortega, 2017; Yang, 2016).

Given that Study 2 examines general proficiency, the remaining section outlines in more detail, studies which have investigated a similar outcome. The key studies which have used C-tests to measure general proficiency in the field are Coleman (1996), Rees and Klapper (2007), Juan-Garau, Salazar-Noguera and Prieto-Arranz (2014) and Hessel (2016).

Coleman (1996) undertook the European Language Proficiency Survey, which served as a large scale, cross-sectional, between-subject design of approximately 25,000 British and Irish language students. Proficiency was measured using C-Tests and administered at each year-group level. The results found mean proficiency scores of students at post year abroad (4th year) to be, on average, 16.36% higher than scores achieved at the onset of their 2nd year of study. No considerable mean gain was found during the first year of domestic instruction, leading Coleman (1996) to conclude that sojourning served as one of the primary variables in accounting for the change.

While Coleman (1996) used a cross-sectional design, Rees and Klapper (2007) undertook a longitudinal, within-subjects design, where the same individual was studied across all four years of their undergraduate study. Fifty-seven British domiciled learners of German were tested on their overall proficiency, as operationalised by C-tests, at four timepoints across a four-year degree. Measures were taken at the onset of the degree programme, at the end of year two (pre-year abroad), at the start of year four (post year abroad) and lastly, at the end of their degree programme. The findings indicated a positive sojourn effect on language development, with the average individual gaining by 8.88 ($sd = 6.89$) marks (9.47%) during the 12 months abroad. There were, however, substantial differences in change scores observed with scores ranging from -6 points to a gain of +26 points. When comparing learning contexts, Rees and Klapper (2007) noted that individuals displayed significantly greater improvement while abroad compared with at-home. It is noticeable, however, that the gain score percentage in this study (9.47%) is lower than that of Coleman (1996) above (16.36%). This difference is likely because Coleman (1996) took his measurement up to a year earlier than Rees and Klapper (2007). Consequently, Coleman's figure includes two-years' worth of growth compared to only one.

Juan-Garau et al. (2014) as part of the Study Abroad and Language Acquisition (henceforth SALA) project investigated changes in overall proficiency in 57 trilingual learners who were taking part in a compulsory three-month ERASMUS period abroad at a UK university. Using a Cloze Test as opposed to a C-test (both types of redundancy testing), learner's proficiency was measured at three time-points, within a within-subjects study design. Baseline (time one) occurred at pre-entry to university, time two occurred 80 hours of L2 instruction at the home university/pre-study abroad (six months), and time three occurred upon return from the UK/post-study abroad. Between timepoint one and two, the average individual demonstrated a non-significant change during the two terms of formal domestic instruction. This result was, however, preceded by a significant ($p = <.001$) increase in proficiency while abroad, with the average learner increasing their score by 9.89%. Comparison of gain scores between learning contexts indicated score change while abroad to be significantly larger than those made at-home.

Hessel (2016) also investigated overall proficiency change as a result of a study abroad. The study undertook a between-subjects design, comparing 99 ERASMUS sojourners and 44 individuals who were either unsuccessful or withdrew their ERASMUS application and of who remained at-home. Data were collected across three timepoints over one year, pre-sojourn, mid-sojourn and post-sojourn. Change scores in the first three months were found to be significant for both the

short-term ($p < .001$, $d = .60$) and long-term ($p = < .001$, $d = .55$) sojourners, but not for those who remained at-home ($p = .08$, $d = .012$). Regression analysis indicated that learning context significantly predicted language change, after controlling for a series of learner-internal and learner-external variables. In the following six months, the long-term group continued to demonstrate significant change, although the extent of this change was not so substantial as the first three months ($p = < .001$, $d = .45$). Similarly, the comparison group also displayed significant change ($p = < .05$, $d = .17$), with the regression analysis indicating that learning context did not significantly predict language gain in the second half of the year.

3.2 Variables Impacting Linguistic Gain

The study of individual differences has been widely explored in recent decades, and there is today a consensus that individuals differ both in their rate of development and in their ultimate level of acquisition. As Kinginger (2011, p. 58) notes: “study abroad intensifies individual differences in achievement: certain students thrive while others founder”. There is today a large body of literature which explores factors associated with individual differences in learner achievement on a year abroad, an overview of which is given below.

Firstly, pre-programme proficiency has long been thought to be the primary factor in accounting for linguistic gain abroad. More proficient users tend to be more confident users, and this is reflected in the amount and complexity of the L2 learners engage in. Higher proficiency levels, according to Dewey et al. (2014, p. 40), enable “deeper connections within the host community.” It has therefore been hypothesised, that because these learners will likely engage in more L2 contact, their gains should be higher (e.g., Dewaele, Comanaru, & Faraco, 2015). As noted by van Niejenhuis, Otten and Flache (2018) however, while high proficiency of the L2 may facilitate language contact, it does not necessarily initiate it.

However, literature has tended to find that, overall, low proficient learners benefit most from a period abroad (e.g., Baker–Smemoe, et al., 2014; Coleman, 1996; Llanes & Muñoz, 2013; Vande Berg, Connor-Linton & Paige, 2009). This finding is perhaps best summed by Freed (1995, p. 27), who states “significant changes do not take place within the study-abroad context at least for more advanced learners, and students make the greatest gains in an immersion environment with initially lower language proficiency.” DeKeyser (2007) argues, however, that these results must be interpreted with caution. After all, the quick acquisition seen in lower proficient learners is likely a result of lexical chunking phrases required to be communicatively competent and not

that of automated knowledge. Moreover, where change scores have been used, this finding may be caused by methodological artefact. As Allison (1990) notes, in tests where scores are finite/bounded, those who score high at baseline have less to gain before reaching saturation, a phenomenon known as the *natural regression to the mean*. Consequently, advanced learner will likely show little change, whereas beginners are more likely to show substantial change.

Highly proficient learners also tend to seek different mediums of interaction. Segalowitz and Freed (2004) and Dewaele, Comanaru and Faraco (2015) found proficient learners to develop in their receptive skills for they were more inclined to listen and watch L2 material (e.g., radio and TV). Studies have typically explored active skills (e.g., speaking), meaning understanding the full potential of study abroad on proficient learners is, to date, not yet fully realised.

Secondly, given the increasing popularity of short exchange programmes in America, it is perhaps surprising only limited literature exists on understanding the interplay between the length of stay and linguistic gain (Llanes, 2011; Yang, 2016). In early studies (e.g., Carroll, 1967; Gomes da Costa, Smith & Whitely, 1975), strong correlations were found in reading comprehension, with longer stays abroad being associated with larger gains. These claims were supported by Davidson (2010), who expanded this relationship to include speaking and listening. Scholars (e.g., Félix-Brasdefer, 2004; Xu, Case, & Wang, 2009) investigating pragmatic development, have also shown that native-like behaviours, develop more consistently, the longer a learner is abroad. Conversely, studies exploring global proficiency have demonstrated little benefit of longer-stays. For example, Rees and Klapper (2007, p. 350) demonstrated “very little advantage in proficiency terms for those who stay for twelve as opposed to six months”. Moreover, Hessel (2016) found that sojourners showed growth in the first three months, and then plateaued over the next six months. Regarding the domains of fluency and accuracy in spoken and written prose, literature has tended to indicate that fluency develops at a faster rate than accuracy with studies predominantly finding accuracy to benefit most from longer stays abroad (e.g., Lara, Mora, & Pérez-Vidal, 2015; Serrano et al., 2012; Storch, 2009).

Thirdly, scholars have long questioned the 'immersive' environment purported by proponents of study abroad programmes, and much attention has been placed on understanding the interactional patterns of sojourners (e.g., Coleman, 2013; Ranta & Meckelborg, 2013; Tragant, 2012; Wilkinson, 1998). While SLA theory would purport the advantageous nature of immersion, in countries where English is used as a Lingua Franca, L1 English speakers may find it difficult

to consistently use their L2 without interference from English (e.g., Dewey, Bown, & Eggert, 2012). Coleman (2015) portrays this idea through concentric circles in which the inner circle comprises of the L2 learner, while the outer circle represents the host community. Consequently, sojourners must show great motivation and effort to expand their social circles so that they may interact with non-L1 speaking individuals. Nonetheless, once access to L2 native speakers is achieved, there is no guarantee that they will want to speak their native tongue, and instead choose to speak English. The literature suggests that social networking and language contact is one of the most significant predictors of language success abroad (e.g., Baker-Smemoe, Dewey, Bown, & Martinsen, 2014; Dewey et al., 2012; Isabeli-García, 2006; Mitchell et al., 2017a; Trentman, 2013). In a number of studies (e.g., Hernández, 2010; Isabeli, 2001), those who had complex social networks with native speakers saw the biggest gains in proficiency and accuracy. Both Baker-Smemoe et al. (2014) and Dewey, Belnap and Hillstrom (2013) observed that individuals who made the largest gains had close bonds with L2 native speakers who were proficient L1 speakers. It was assumed that by having these friends, integration into the host community was aided and allowed sojourners to have direct contact with members of the host community who may otherwise have been hostile. Magnan and Back (2007), on the other hand, found that the amount of L2 contact did not correlate with linguistic gain, instead finding prior coursework to be the highest correlate.

The extent of and types of interaction among learners is also explored in the literature. Mitchell et al. (2017a) for example, found that while learners used the L2 regularly, many failed to live up to the ideals they had placed on themselves before departing, serving as an area of regret for many. While small talk and service encounters occurred in the L2 almost daily, activities such as internet browsing, and text messaging were often conducted in the L1. Across time, contact with the L2 generally depreciated. Moreover, activities such as reading academic texts, reading literature and watching TV were found to take place infrequently in both languages.

Fourthly, the ‘where’ and the ‘who with’ has long been considered as a factor in accounting for individual differences. Given the propensity of American-orientated studies, homestays have frequently been studied. Diao, Freed, Smith and Khawaja (2011, p. 128) found homestays to be an “overwhelmingly valuable experience for both linguistic and cultural development”, finding sojourners to view their homestay families as a second family and being forced to speak the L2. Kinginger (2014, p. 54) on the other hand, argued that “the homestay is not a reliable environment for language learning abroad”, and that “no absolute correlation has been found

between living arrangements and the development of proficiency” (p. 54). This view is supported by Schmidt-Rinehart and Knight (2004), who noted that while sojourners viewed homestays as a positive experience, they did not report spending any considerable time interacting in the L2.

From a European perspective, homestays are relatively rare, and instead, sojourners tend to go into either university halls or private rented accommodation. Milton and Meara (1995) found that those who stayed in private accommodation made greater progress than those who lived in university halls. This finding was supported by Klapper and Rees (2012) who found ‘gainers’ stayed in private accommodation, often with L2 native speakers; whereas ‘non-gainers’ resided in university halls with other international students where English served as a *Lingua Franca*.

Lastly, only a handful of studies have discussed language gain in relation to placement type (i.e., sojourner role), and even fewer have made a specific link between change scores and sojourner role. Willis, Doble, Sankarayya and Smithers (1977) found those on a work placement to significantly improve. They compared work placement students with a small group who studied at a university, finding that those who worked outperformed those who studied on all linguistic measures. They hypothesised this result was because workers were afforded more opportunity to interact in the L2. Meara (1994) explored perceived change collecting survey responses from 586 participants (301 on student exchange, 129 teaching assistants, and 81 workplace internships). While all groups noted how they felt the year abroad was beneficial, the university group perceived their linguistic growth to be smaller than that of the other two groups.

3.3 Personality Within the Field of Second Language Acquisition

Personality and the field of SLA have shared a complex relationship. At the core of SLA research, is the ability to explain considerable variability which exists between individuals in the rate and outcome of acquisition success. As aforementioned, despite being in similar environments, individuals differ to the extent in which they master an L2. Consequently, it may be that psychological sources can account for this variability and their subsequent interplay with the learning environment (e.g., Dewaele, 2009; van Daele et al., 2006).

There remains a perception that successful language learners carry with them a unique set of personality traits which can aid acquisition originating with a study conducted by Lalonde, Lee and Gardner (1987). Here, they found that 83% of teachers perceived good language learners to have 11 trait behaviours, including being sociable, flexible and imaginative. However, since this

publication, personality has garnered little attention in the literature despite its assumed influence (Dörnyei & Ryan, 2015).

This level of reluctance may be borne out of several factors. Firstly, by incorporating personality into one's research, one must take a cross-disciplinary approach. Often, SLA researchers lack such expertise, with further study required to understand the theoretical frameworks of personality together with the appropriate research skills (Dewaele, 2009). Given, the time required to learn such skills, scholars have tended to focus solely on their area of expertise. Secondly, disentangling personality from the multitude of cognitive, social and environmental factors that contribute to SLA can be challenging. Some traits remain invisible in some tasks while surfacing in others and as such, trying to isolate the specificity of personality of SLA is difficult (Dewaele, 2012). Lastly, selecting the appropriate measure and instrument can prove to be a complex task. There is today a multitude of personality and linguistic measures, each which tap into different aspects of the outcome measured and are designed to be used in a particular manner. Scholars can be discouraged to use psychometric instruments which may appear foreign to them, therefore choosing to avoid the construct they measure (Dörnyei & Ryan, 2015).

3.3.1 Second Language Acquisition and the broad traits

Within the SLA field, personality has often been viewed within the broad trait framework (e.g., Five-Factor Model), with the following section providing an overview of research for each trait.

Openness to Experience has consistently been found to be a strong predictor of language success, particularly with regards to study abroad programmes. Verhoeven and Vermeer (2002) were one of the first researchers to establish a link, finding openness to aid the development of basic pragmatic, lexical and syntactic skills among young L2 Dutch learners. This finding is further supported by Ehrman (2008) and Ożańska-Ponikwia (2010), who found learners who scored high in openness acquired native-like ways of self-expression and felt different when using their L2. All three findings indicate a strong relationship between openness and pragmatic ability.

Very little evidence is available, which directly measures the role trait conscientiousness plays in language learning. Ehrman (2008) describes these individuals as always wanting to improve and use a variety of strategies in order to improve, including goal setting and self-assessment. She consequently hypothesises that conscientious learners should be more successful in acquiring an

L2. Wilson (2008) indicated that conscientious learners held greater motivation towards their studies, finding these learners to be less likely to quit a language course. Such evidence tentatively concludes that conscientiousness can have an indirect effect on L2 acquisition.

Of the five traits, extraversion is the most studied (van Daele, Housen, Pierrard, & Debruyne, 2006) with SLA theory purporting that extraverted learners should be stronger language learners for they are more inclined to engage actively with the L2 (e.g., Krashen, 1985; Swain, 1985). However, evidence remains inconclusive regarding the relationship between *oral* production and extraversion, with studies typically being correlational. Dewaele and Furnham (2000), for example, found a positive correlation between extraversion scores and utterance length, number of filled pauses and speech rate. Van Daele et al. (2006), on the other hand, found development in oral fluency, complexity and accuracy to be unaffected by extraversion scores over 12 months as demonstrated by non-significant interaction effects. Concerning writing, Robinson, Gabriel and Katchan (1994) found introverts outperformed extroverted learners in written tests, whereas extroverts tended to perform better in oral measures. In sum, the impact of extraversion appears to be dependent on the language variable under-study with causal inference weakened due to the lack of longitudinal studies.

Although extremely limited, research would suggest that agreeableness is linked with the development of L2 pragmatic competencies. Ożańska-Ponikwia (2010) found that when interacting, agreeable L2 learners were more likely to notice their own facial and other non-verbal expressions. Within the context of study abroad, it may be hypothesised that agreeable behaviours may aid integration and with it, afford learners more opportunities to communicate in the L2. MacIntyre and Charos (1996, p. 19) for example, found a link between agreeableness and willingness to communicate, stating “people who are more pleasant and agreeable themselves would be more likely to have pleasant contacts with target language group members, and this appears to be reflected in their willingness to communicate.” Similar to that of openness, agreeableness appears to facilitate pragmatic acquisition and given that agreeableness can be linked to acculturation, appears to have an indirect effect on acquisition through means of affording learners greater L2 opportunities.

Neuroticism at the trait level has received little attention in SLA research (Dewaele, 2013), although the available literature indicates that neurotic individuals tend to be successful learners. Robinson et al. (1994) found highly neurotic learners to outperform their peers on both oral and

written tests, while Ożańska-Ponikwia (2010) found neurotic individuals to have fewer difficulties in expressing their emotions both in their first and second languages. Although perhaps counterintuitive, it would appear that neuroticism can facilitate L2 acquisition.

To conclude this section, finding a direct link between language gain and the broad traits has not been forthcoming in the literature. Scholars have instead tended to subsume the broader traits within more defined constructs (e.g., willingness to communicate (WTC) and extraversion), choosing not to make direct links between the broader traits and language change.

3.3.2 Second Language Acquisition and the narrow traits

As noted by Dörnyei and Ryan (2015), many of the meaningful findings have emerged from researching personality at the lower-order trait level. This is an exciting research direction but calls into question how useful broad trait models are in our understanding of personality in SLA.

Anxiety has received by far the most attention of all behavioural characteristics, thanks in part to the development of the 'Foreign Language Classroom anxiety' instrument by Horwitz, Horwitz and Cope (1986). Anxiety is generally assumed to be negatively associated with L2 achievement (Teimouri, Goetze, & Plonsky, 2019). Of the four skills, anxiety has typically been found to be most detrimental to speaking, with highly anxious learners experiencing a greater number of breakdowns in speech runs and actively seek less L2 interaction (Ożańska-Ponikwia, 2010, Dewaele, 2009). Anxiety has also been found to harm listening comprehension (Elkhafafi, 2005), reading (Sellars, 2000), and writing (Cheng, 2002). Moreover, anxious learners consider themselves to have lower perceived self-worth (Dewaele, 2010) and are more likely to want to discontinue studying foreign languages (Dewaele & Thirtle, 2009).

On the other hand, curiosity and resilience have received little attention to date, and have been viewed through the lenses of other SLA constructs (e.g., learning strategies, language motivation) as opposed to separate entities.

The construct of curiosity has typically been dichotomised between that of *L2 communicative curiosity* and *L2 linguistic curiosity* (Mahmoodzadeh & Khajavy, 2018). Strong *communicative curiosity* is related to one's desire to communicate in the L2, while *L2 linguistic curiosity* is associated with one's drive to overcome gaps in their L2 knowledge. Having high levels of each are considered to facilitate acquisition and can be linked to the constructs of Willingness to Communicate and

language motivation, both variables long cited as accounting for individual differences in learning success (Dörnyei & Ryan, 2015; Ellis, 1994). There is no literature, to the best of my knowledge, which makes a direct association between curiosity and language development.

Developing resilience in language learners has been seen as a way of encouraging long-term learning (MacIntyre & Mercer, 2014) and learner autonomy (Puppel, 2012). Through developing resilient language learners, individuals will be tolerant of their own mistakes, be adaptable in the classroom, and can self-reward themselves (Puppel, 2012). Again, such characteristics are assumed to facilitate language acquisition, but to the best of my knowledge, no study has explicitly studied the link.

3.4 Personality, Second Language Acquisition and Study Abroad

When investigating the role personality plays in language acquisition, specifically within the context of study abroad, little evidence exists. To the best of my knowledge, only one study (Baker-Smemoe et al., 2014) has explored objective language change in sojourners using a trait personality framework.

Baker-Smemoe et al. (2014) examined many predictors, including personality, to establish if any accounted for linguistic change witnessed. The study captured data from 102 American sojourners, who were abroad for between eight and 16 weeks. Pre-programme proficiency ranged from novice to advanced, while the average age of the participants was between 21.94 and 24.7. Competence was measured using the ACTFL Oral Proficiency Interview administered at the beginning and end of the learner's sojourner programme. The NEO Five-Factor Inventory was also administered at the same time. Upon return home, learners were split between 'gainers' and 'non-gainers' and a series of statistical analysis conducted to establish whether any significant differences in the predictors existed. For personality, no significant differences between the groups were found although 'gainers' did score higher in extraversion, openness and conscientiousness, and lower in neuroticism. The study found that only social networks, pre-programme proficiency, and pre-departure intercultural sensitivity were significantly different in 'gainers' and 'non-gainers', suggesting these were the significant predictors of language success on a sojourn.

Studies measuring language change through self-reported measures have also shown interest in understanding how personality differences may account for individual differences. Given the

self-report nature of the language measure, the findings can be considered less robust than Baker-Smemoe et al. (2014) but will nonetheless be explored below.

Arvidsson et al. (2018) investigated perceived linguistic change in 59 Swedish and Belgian undergraduate sojourners, correlating such change with personality change. All sojourners spent between three and five months abroad. The personality measure was administered pre/post-test, while returners were also asked to report time spent speaking the L2 in an average week and perceived linguistic change in speaking only. The results suggested a significant correlation between self-perceived progress in speaking and Cultural Empathy (closely linked to openness) ($r = 0.30$; $p = <.05$). No other trait significantly correlated with perceived growth, suggesting that those who spoke the L2 more often saw a greater trait manifestation in Cultural Empathy only. Significant correlations were also found between the amount of spoken target language per week and Cultural Empathy ($r=0.31$, $p<0.05$) and Open-mindedness (again closely linked to openness) ($r=0.28$, $p<0.05$). Therefore, those who spoke more L2 outside the classroom became more sympathetic towards the host community and were more willing to try new things.

Van Niejenhaus et al. (2018) investigated the link between language proficiency and cultural integration. The study based on 163 sojourners ($n = 163$) completed a self-report linguistic measure and the Multicultural Personality Questionnaire at two timepoints across three months. Self-report questions included “How well do you write Dutch?”, on a 7-point Likert Scale. The results showed that positive perceived linguistic change was predicted by two constructs of the MPQ; identification to the host culture and attitudes towards the host culture.

Chapter 3 has examined second language acquisition, a construct explored in both Study 1 and Study 2. An overview of the thesis’s research questions is now presented.

3.5 An Overview of the Thesis’s Research Questions

This section provides an overview of the research questions this thesis addresses. As presented in section 1.8, the thesis is divided into two studies. Study 1 serves as a systematic review, with the purpose of examining:

RQ1: How effective are study abroad programmes in achieving their fundamental aim of improving linguistic ability compared to those who remain in the domestic classroom?

Study 2 addresses a number of subsidiary research questions, as opposed to an overarching research question as per above. These questions focus on the constructs described in Chapter 2 (personality and well-being) and Chapter 3 (Second Language Acquisition) and are as follows:

RQ2: Do sojourners experience significantly greater personal growth as measured by broad and narrow personality traits compared to non-sojourners?

RQ3: Do sojourners experience significant broad trait personality change over time, and is this uniform across all sojourners?

RQ4: What are the predictors of broad trait change?

RQ5: What is the breakdown in variability at the three levels of personality?

RQ6: How is variability in state agreeableness related to individuals' perception of situational characteristics?

RQ7: Do sojourners experience significantly greater well-being over time compared to non-sojourners?

RQ8: Do sojourners experience significantly higher well-being over time, and is this uniform across all sojourners?

RQ9: Are monthly well-being scores contingent on monthly narrow personality scores?

RQ10: What is the relationship between average well-being scores and the potential predictors?

RQ11: Do learners return home with higher linguistic proficiency after a year abroad?

RQ12: Does personality serve as a valid individual difference in linguistic gains made?

Research questions 2 and 7 serve as efficacy questions and look to capture whether one group changes differently to another group over time, based on a differing factor (i.e., learning context). These research questions compare data collected from both sojourners and non-sojourners.

Research questions 3, 4, 8, 9 and 10 focus on the sojourn sample and the reasons for doing so was twofold. Firstly, the overarching aim of the research project was to inform study abroad practitioners and policymakers of how the study abroad may shape the personality and well-being of those who undertake it. Secondly, there exists a larger evidence base concerning how the variables of interest change in an at-home environment. For example, from a well-being perspective, scholars (e.g., Bewick et al., 2010) have conducted longitudinal studies capturing how well-being changes across tertiary study. While these research questions do not infer causality, given the lack of control group, they can provide a descriptive account of variable change.

Research questions 5 and 6 investigate the whole sample, regardless of learning context. The rationale behind this is that findings are not learning context dependent. These questions do not explore whether change is attributable to a particular learning context, nor are findings of interest generalisable to a defined audience (i.e., study abroad practitioners).

Research questions 11 and 12 focus on the sojourners who were partaking in a Modern Foreign Language degree specifically. Given that no at-home students were Modern Language majors, it was only possible to answer these questions with the sojourn sample.

Chapter 4 presents the first of the two studies which makes up this thesis, with Study 1 serving as a systematic review which can place itself within the field of Second Language Acquisition.

Chapter 4: Study 1 – Value of Study Abroad on Linguistic Gain; a Systematic Review

As outlined in the introduction, this PhD comprises of two, distinct, but complementary studies: Study One – described in this chapter is a systematic review into the effectiveness of a study abroad intervention on a series of linguistic outcomes. Study Two, as described in the next chapter, situates language development within the context of the ERASMUS programme and establishes the efficacy of sojourning on personal development compared to remaining at-home. The two studies, therefore, work in unison, with the systematic review complementing the empirical component of the research in order to answer the broader research question of the PhD thesis; specifically, does study abroad serve as a valuable intervention in facilitating linguistic and personal growth?

4.1 Introduction

Each year, thousands of language students worldwide undertake a study abroad programme. This programme serves as a natural hiatus from regular studies in which learners can immerse themselves in a foreign culture and/or language. For this chapter, study abroad is defined as the following, taken from Kinginger (2009, p.11):

a temporary sojourn of pre-defined duration, undertaken for educational purposes

The role of learning context in SLA has garnered much attention in recent decades, mirroring the increasing globalisation of education (Jackson, 2018). The learning contexts under study can be divided between study abroad, immersion, and at-home, all of which differ in the intensity of linguistic exposure and opportunities afforded to the L2 learner. Traditionally, research has focused on the learning contexts of study abroad and at-home in order to infer the extent to which language develops abroad (Grey, 2018).

This explosion of interest is reflected by the sheer volume of publications exploring language gain abroad, reaching a historical peak between 2011 and 2014. Since 2014, several journals, including 'System', have published special issues on the topic, while in 2016, the creation of the journal 'Study Abroad Research in Second Language Acquisition and International Education' was devoted explicitly to language learning abroad. Therefore, it can be said that the domain of study abroad research has today reached maturity (Tulloch & Ortega, 2017).

Historically, publications have tended to concern themselves with an American undergraduate population (Yang, 2016). While there is no definitive reason for this, factors at play potentially include: 1) American-domiciled journals (e.g., 'Frontiers') showing interest in study abroad research (Rees & Klapper, 2007); 2) shorter lengths of stay resulting in swifter evaluations of interventions (Kinginger, 2009); and 3) the difficulties faced by international students regarding gaining visas for entry into English-speaking countries (Yang, 2016). This skew is surprising given that, within a European context, the European Commission administers the ERASMUS programme, which is, to date, the largest exchange programme in the world. Since its inauguration, the ERASMUS programme has allowed over three million people to study or train abroad, and a primary driver of this has been the desire to improve linguistic proficiency (British Council, 2005; Teichler, 1997). This imbalance towards American-based literature is beginning to be addressed by scholars, with several European-based language projects recently being completed. Examples of these include the LANGSNAP project in the UK (see Mitchell et al., 2017a), the SALA project in Spain (see Pérez-Vidal, 2014), and the SAREP project (see Howard, 2019) in Ireland, due for completion in 2020. These studies offer an insight into the developmental patterns of European sojourners, who can differ in several factors from their American counterparts. These differences include length of stay, starting proficiency and psycholinguistic variables (see Coleman 1996). The readdressing of this imbalance is an important step in the field of study abroad research and with it, hopefully, provide more insightful findings across the varying exchange programmes available.

This growing body of literature has also further highlighted the disparity between public consensus and empirical evidence. While it may be assumed that study abroad serves as a natural path to language learning success (Barquin, 2012; Hessel, 2016), this view has not always been reflected in the literature. Overall, empirical research has shown that developments are subtle, vary considerably across individuals and are dependent on the linguistic competency measured. If expectations concerning language development pre-sojourn are not matched post-sojourn, learners will likely have a sense of disappointment and non-progression (DeKeyser, 2007; Härkönen & Dervin, 2016; Wilkinson, 1998). This disappointment can not only impact on their immediate undergraduate programme but also in later life (e.g., career decisions) if individuals become disillusioned with the second language (Yen & Stevens, 2004). There does, therefore, need to be clear documentation for study abroad practitioners to refer to and share to ensure expectations are matched, which will ensure that learners can fully benefit from the sojourn experience.

Further difficulties in affirming conclusions regarding development arise from the methodological limitations found in the literature. As noted in early work by Freed (1995), studies which do not include a control group are of limited value, for, without such comparison, changes seen in the treatment group cannot be causally attributed to the intervention itself. It is unknown whether learners would have made a similar gain in domestic formal instruction.

In summary, despite the extensive body of evidence available, scholars have continually found it difficult to substantiate the assumed linguistic benefits of a sojourn experience and perhaps more pertinently ascribe witnessed change to the study abroad experience itself. This is perhaps best summed by Kinginger (2009, p. 213) who noted a decade ago that the field of study abroad research is “a largely uncoordinated and piecemeal affair, with individual scholars or small groups of researchers pursuing diverse interests within their own institutions.”

Before embarking on the current review, a scoping review was conducted, revealing one existing meta-analysis (Yang, 2016) and a scoping review conducted by Tullock and Ortega (2017).

Yang (2016) investigated the effectiveness of study abroad on L2 development, together with understanding the role length of stay plays in facilitating this development. In total, 66 studies were found to match the inclusion criteria, which were then coded based on theoretical and methodological characteristics. Such characteristics included research methodology and target language outcomes (e.g., oral). From these 66 studies, 11 were identified as being eligible for the meta-analysis. The inclusion criteria for eligibility included studies requiring a control-group design, and the dependent variable had to involve a measurement of participants' linguistic proficiency. During the analysis, Yang found 65 effect sizes from the 11 studies, of which were then averaged per study, so that each study had one effect size, for comparability purposes. Post-test comparisons of the 11 studies showed effect sizes ranging between 0.5 and 7.8, with a mean weighted effect size of 0.75, suggesting a large-sized effect overall. Yang (2016, p. 78) notes that these data support the view that studying abroad “could lead to greater L2 linguistic attainment compared to AH classroom learning.” Of the studies included, only one was conducted within a European context, while no study post-2011 was included. Moreover, the inclusion criteria provided no requirement for a study to have baseline equivalence. Reporting baseline equivalence is important because if achieved, one can be more certain that any differences in the outcome variable are as a result of the treatment itself. If unreported, or uncontrolled for, results may be skewed because the treatment group were already systematically different from the control group before undertaking the intervention (Christensen, Johnson, & Turner, 2014). For

example, one study in this meta-analysis had an effect size of 7.797, yet at baseline, the sojourn group was significantly more proficient in several variables.

Tullock and Ortega (2017) carried out a scoping review of the SLA literature, with the purpose of evaluating the literature regarding whether sojourning facilitates the acquisition of oral fluency. The authors explored two databases (ProQuest and ERIC), together with grey literature (Georgetown University library website and Google Books), yielding a total of 408 hits after duplication. This searching process was iteratively run for three years (2014-2017). The final database consisted of 401 publications. Of these, 31 oral-fluency outcome-orientated studies were identified and of which meta-analytic methods were applied. Tullock and Ortega (2017) noted that the characteristics of the included studies were as follows:

- Primarily focused on two populations: English L1 speakers and English L2 speakers.
- Majority of studies explored a length of stay of one semester.
- Lack of control for participants initial level of L2 proficiency; this was either unreported or uninterpretable on information provided.

Regarding the primary research question “*can clear SA benefits for oral fluency be claimed?*”, the results of the meta-analysis proved inconclusive. Several domains found positive results (e.g., speech rate, articulation rate, pause length), while other domains were less conclusive (e.g., mean length of run, pause frequency) demonstrating positive, null and negative effects. Consequently, Tullock and Ortega (2017, p. 14) concluded by stating “in sum, we can answer only tentatively, that students probably become more fluent after a SA experience.” In comparison to this current systematic review, Tullock and Ortega (2017) explored changes in the sojourn group only. Consequently, their scoping review did not evaluate whether sojourning facilitated the acquisition of oral fluency over that of formal domestic instruction.

In this systematic review, all available studies which met the inclusion/exclusion criteria were included. Included studies were not exclusive to a particular region or exchange programme, nor language outcome. The review takes advantage of the most recent literature concerning language change while abroad, including studies within a European perspective and examines more than one linguistic construct systematically.

4.2 Design and Methodology

Systematic reviews can be defined as “the application of strategies that limit bias in the assembly, critical appraisal and synthesis of all relevant studies on a given topic” (Chalmers, Hedges, & Cooper, 2002, p. 17). Traditionally, these reviews were found in the medical field, where their usage was designed to evaluate health care practices. Their usage has, however, been ever-growing in the field of Education and Social Sciences in the past several decades (Torgerson, 2003; Clarke & Chalmers, 2018). The need for this type of review stems from the ever-growing body of literature, which can mean that many publications receive little to no attention. It is the purpose of these reviews to search for, evaluate and synthesise the literature so that results can be easily read and understood outside the prism of academia (Gough, Oliver, & Thomas, 2017).

Systematic reviews differ from traditional literature reviews in that their methods are designed to be explicit, transparent, replicable and open to scrutiny (Torgerson, 2003; Torgerson, Hall, & Light, 2012). While literature/narrative reviews may summarise studies, there is no rationale given as to why a certain study may have been included or excluded, and perhaps more pertinently, potentially relevant and important studies may have been missed from the review. However, because systematic reviews present results which been formed through explicit, rigorous and accountable methods, the reader can access the appropriateness of the conclusions presented by the review (Gough et al., 2017).

Typically, systematic reviews can be thought of as a seven-stage process (Cooper, 2017). First, one must define the problem and formulate the research question. Traditionally, reviews have been associated with, although not limited to, research questions concerning efficacy which looks to evaluate the impact of a policy or treatment characteristics on one or more outcomes. The protocol is also developed during this stage, which outlines the plan of the systematic review, and the theoretical considerations underpinning the research question. Secondly, the literature search is undertaken, with this being followed by a coding stage. Here, reviewers must code each piece of relevant information to the research question. This can include study characteristics, research design and relevant findings. Next, the literature must be screened based on a pre-determined inclusion criteria and coding system. Once all studies have gone through several rounds of screening, the outcomes of all included studies must be analysed. If statistical data is collected, a meta-analysis will be typically produced. This may, however, be substituted by a narrative synthesis if a meta-analysis is deemed unwarranted. Next, quality appraisal of the included literature should be undertaken in order to ascertain whether the conclusions presented

are fair and warranted. Interpretation can also be offered to the generalisability of a review's findings. Lastly, the findings of the review should be made available in a publicly accessible document.

Systematic review design can be considered to have a number of strengths over the traditional literature review design. Firstly, systematic reviews aim to be open and transparent. In being so, future replication is made easier and more feasible (Cooper, Hedges, & Valentine, 2009). Secondly, systematic reviews aim to minimise bias in each of the stages presented above. For example, systematic reviews aim to seek out all available literature on a given topic, ensuring a wider breadth of research is captured than the traditional literature review, where papers may be included or excluded without reasoning. Whether all relevant literature can be fully discovered is open to debate, however, and the point in which saturation is achieved may differ from researcher to researcher (Cooper, 1987). Moreover, as reasons for study exclusion are made explicit at each stage of the review process, readers can make their own judgement over the appropriateness of this decision. As noted by Torgerson (2003, p. 6), this degree of openness can mean that findings are often "less susceptible to selection, publication and other biases" which may exist in other types of reviews. Lastly, by synthesising findings across multiple studies, findings are often more generalisable, consistent and precise across a sub-group of individuals (Centre for Reviews and Dissemination, 2009). Furthermore, synthesising evidence in a clear and structured way allows for policymakers and practitioners to make quicker, more evidence-based decisions on best practice and ensure that time, money and labour are not invested into an intervention which does not work.

Conversely, systematic reviews have been criticised by some who see the narrow sample of studies included in the review as ignoring much of the relevant available research. It is the purpose of the author to make clear throughout the review, the rationale behind selecting these studies, and ensure that the conclusions presented are a fair representation of the studies included (Gough et al., 2017). This is essential if one is not to misrepresent or bias their conclusions in any way. Secondly, while researchers aim to seek out all available literature, this is very much dependent on the resources available to them. For example, there are monetary costs to financing online databases, and different institutions may have differing levels of access to these databases. This can introduce bias into the searching process and may mean that a potentially high number of important studies may be missed (Mallett, Hagen-Zanker, Slater, & Duvendack, 2012). Linked to this point is the fact that the funding of systematic reviews may

result in a conflict of interest. Donors may be associated with a particular intervention and as such may undermine the objectivity of results, particularly if findings go against the opinions of the donors themselves. Lastly, in review teams which involve large numbers of staff, interpretation of the inclusion/exclusion criteria is likely to be different among reviewers. While disagreements can be mediated via discussion or piloting implemented to ascertain disagreement rates, there will always be a level of subjectivity in how the criteria are interpreted (Mallett et al., 2012), potentially limiting replicability.

4.2.1 Research question

The purpose of the systematic review was to provide a summary and synthesis of the literature in the topic area of study abroad. All methods are explicit, transparent and replicable, and strategies were employed to minimise the risk of bias and random error. The design and methods were informed by several methodological works including the ‘Cochrane Collaboration Handbook (Higgins & Green, 2008); An Introduction to Systematic Reviews (Gough et al., 2017), ‘Systematic Reviews’ (Torgerson, 2003) and ‘The Handbook of Research Synthesis and Meta-Analysis (Cooper, Hedges, & Valentine, 2009). Before conducting the review, a protocol was developed, in which the design and methods were outlined, together with inclusion/exclusion criteria (Table 3) and a protocol (Appendix A).

This systematic review aimed to collate all available literature which adequately assesses the effectiveness of study abroad in facilitating linguistic gain. The research question for the systematic review was:

RQ1: How effective are study abroad programmes in achieving their fundamental aim of improving linguistic ability compared to those who remain in the domestic classroom?

4.2.2 The importance of study design

The research question looks to evaluate the intervention of study abroad as a means of facilitating the development of linguistic ability. Within any evaluative question, the notion of causality is vital. In order to successfully evaluate if an intervention has worked, one must be certain that the effect seen is indeed caused by the treatment itself as opposed to any other possible confounding variable. Study designs differ in the strength of causal inference offered, and this is explored further below.

When conducting effectiveness research, randomised controlled trials (RCT) are considered the 'gold standard' of study design (Torgerson, C., & Torgerson, D., 2003), for they are considered the most robust method in measuring efficacy (Torgerson, 2003). Within an RCT design, two or more groups are randomly allocated to either a treatment or control group and are as such, considered equivalent (with the exception of chance) in both measured and unmeasured variables. Consequently, any outcome differences that are observed at post-test can be considered as a result of the intervention and not due to any baseline differences between the two groups that may have existed prior to the study (Shadish, Cook, & Campbell, 2002). Put differently, RCTs allow for confounding variables to be controlled for and in doing so reduces the plausibility that other alternative explanations can explain the observed outcome.

In many cases, however, particularly in social science, random assignment is not deemed feasible and instead a quasi-experimental design (QED) may be undertaken. The difference between RCTs and QEDs is that the latter lacks random assignment. In QEDs, assignment to conditions is undertaken by either self-selection (i.e., individuals decide themselves which group to go in) or by administrator selection (i.e., a teacher/parent/carer or other decides which group the participant enters into). It is here, where the QEDs become less robust in causal interpretation for the lack of random assignment potentially introduces selection bias into the study, whereby the treatment and control group may differ in systematic ways beyond the variables studied. Consequently, any findings or conclusions made are susceptible to the observed effect being as a result of these systematic differences opposed to the actual treatment.

A commonly used quasi-experimental design is that of the pre-test/post-test control group design. Here individuals are split across two groups and measured prior to and after the intervention. Because the groups are non-equivalent, there will always be a potential for selection bias, but the pre-test measure serves as a vital component in ascertaining how strong this likelihood is (Shadish et al., 2002). The further apart the two groups are at pre-test, the stronger the chance of selection bias. In brief, the strength of causal inference is dependent on the way individuals are assigned to groups, and whether a pre-test accounting for differences at baseline has been used. Regardless, the non-equivalent comparison group design will rarely give findings which are a close approximation to if random assignment had been used (Shadish et al., 2002).

Alternatively, a within-subject design may be implemented whereby all participants receive both the treatment condition and control condition over an extended period. Such a design is popular

in instances where a between-subject design is not possible, such as in the UK based SLA research where sojourning is compulsory for UK language learners (e.g., Rees & Klapper, 2007; Mitchell et al., 2017a). The design has the advantage of requiring fewer participants to achieve the same statistical power as a between-group design and reduces any unsystematic variability for most individual differences are held constant. Therefore, it can be argued that any differences observed over time can be associated directly with the changes in a particular variable (e.g., learning context). The within-subject design is, however, open to criticism, predominantly due to the wide-ranging number of threats to internal validity. Participants may demonstrate improvement over time as they become accustomed to the testing materials (practice effect), or instrument decay whereby a particular instrument changes over time. For example, a spring would likely weaken over time, while a human judgement may be stricter or more lenient over time. Other threats to internal validity include history effects and fatigue effects (Gorard, 2013).

Other experimental designs exist including Time-Series design and Regression Discontinuity design. In an interrupted time-series design, the researcher takes measures both before and after a treatment has been introduced. If a change in level or slope is present after the treatment condition, then this is taken as evidence that an effect exists. The biggest threat to internal validity in regression discontinuity design is that of history. History bias exists when something in addition to the treatment may have occurred at the same time that the intervention was implemented and as such one cannot be sure what has caused the effect (Christensen et al., 2014). Regression discontinuity design refers to a design which looks to establish whether a treatment is effective within a group of individuals, pre-determined by some set criteria. An assignment measure is undertaken, and once selecting a cut-off criterion, some individuals partake in the treatment group, while others enter into the comparison group. The primary threat in this design is a differential history effect, where one group of participants experience something beyond the treatment that the control group does not (Shadish et al., 2002).

Quasi-experimental designs can also be longitudinal or cross-sectional in nature. Longitudinal designs are more akin to the classical experimental design where individuals are measured at two or more timepoints (e.g., pre-test/post-test). Longitudinal designs are more suitable at capturing change in an outcome measure. Cross-sectional studies, on the other hand, measure the outcome variable at a single point in time. Cross-sectional designs are more adept at evaluating variation between participants in the outcome measure at one point in time and cannot capture change. Cross-sectional designs can, however, not prove causality because there is 'ambiguity about the

direction of causal inference'. As there is no time order to the variables, it is unclear whether A has caused B, or whether B has caused A, severely threatening internal validity (Bryman, 2012). Lastly, any study design which does not have a control group design cannot infer causality, nor be used for any research question concerning the evaluation of an intervention programme. Without the formation of a control group, it is impossible to ascertain whether the same individual would not have made the same development while in the comparison group, and it is also impossible to infer that the intervention was the cause of any change seen with any great certainty (Christensen et al., 2014).

In brief, in order to infer causality, and conclude that an invention does indeed influence the outcome being measured, an RCT should be utilised. In instances where an RCT is unfeasible, then a non-equivalent control group design should be implemented for this design does warrant causal inference. Evaluating the strength of an intervention is severely weakened when using a cross-sectional design, while no evaluation should be attempted when no control group exists within the design.

4.2.3 Inclusion/Exclusion criteria

The inclusion/exclusion criteria were set out in the review protocol and can be found in Table 3. These criteria were developed before conducting electronic searching. In order to be included, studies had to be on the topic of linguistic development during a study abroad programme. Participants had to be adult learners (18+) who were completing a period abroad as part of their academic language studies. The sojourn itself had to last for longer than five weeks, excluding studies which looked at summer schools or short-term sojourns as this is often not reflective of the 'study abroad' experience. Outcome measures could have been any component of language learning (e.g., oral, vocabulary, writing) although outcome measures must have been objectively measured as subjective measures have repeatedly been shown to hold little reliability (e.g., Mendelsohn, 2004). No time restrictions were implemented, although any publication post-April 2019 would not have been found when searching. All published and non-published material was included, minimising the threat of publication bias. This is important as the existence of publication bias may result in programmes showing an exaggerated treatment effect.

Table 3: Inclusion/Exclusion Criteria

Included	Excluded
Topic: Study abroad, including affiliated organisations such as the British Council.	Topic: Non-study abroad related interventions
Date: No time restriction	Date: -
Publication status: All published and unpublished material in the public domain	Publication status: N/A
Study design Non-ERASMUS: Any study design where there is a control or comparison group – RCT (individual and cluster); quasi-experiment (interrupted/control time-series designs, control group post-test only, control group pre/post-test). ERASMUS: All of the above AND pre-experimental designs (e.g., pre/post-test with no control group)	Study design Non-ERASMUS: Case-study designs; designs with only post-test and no control group; basic time-series designs. Review articles and non-empirical literature. ERASMUS: Case-study designs. Review articles and non-empirical literature.
Participants: Any undergraduate/postgraduate student undertaking a study abroad as part of their academic degree studies. Control students must be a comparable group (e.g., matched comparisons at baseline)	Participants: Non-academic learners or are under the age of 18.
Intervention: Studies which include a study abroad of which is longer than five weeks in length.	Intervention: Does not have a study abroad component. A length of stay less than five weeks. Summer school programmes.
Outcomes: Studies which learners are measured at post-test on any linguistic skill and their relevant skill outcome, e.g., speaking, writing, reading, listening, pragmatics. This can be measured through multitude instruments – for example, length of utterances, length of prose, speech/written accuracy/fluency, reading score, listening score, grammatical score. The outcome must be objective (i.e., not self-report).	Outcomes: Measures not looking at linguistic gain, e.g., intercultural competency. Outcomes which are self-rated/perceived change, e.g., on a scale of 1 – 10, how much do you believe you have improved?

4.2.4 Study designs

As explained earlier, the systematic review aimed to synthesise studies situated in both a European context and non-European context. It was decided during the early stages of this review that a separate criterion would be created for study design, dependent on whether the study was European or non-European based. The decision to do so was threefold. Firstly, it was thought that if a strict criterion were implemented, no ERASMUS study would be carried forward to the main analysis, given the small pool of available literature. Secondly, the ERASMUS programme offers administrative flexibility which means a suitable control group is not always possible. In the UK, for example, a one-year sojourn programme is compulsory for language learners and as such, no UK-based research study can implement a comparison group design (Mitchell et al., 2017a). In Spain, some study abroad programmes are compulsory for three months, spending time both at-home and abroad within the same year (Pérez-Vidal, 2014), while Germany offers no compulsory programme (Hessel, 2016, 2017). Lastly, by including European-based research, this review assimilates itself with the context of the thesis. Study design criteria are split across two types of exchange programmes.

4.2.4.1 *Non-ERASMUS*

Any study fell into this category if the exchange programme in question was not the ERASMUS programme. Given the weight of literature towards non-ERASMUS exchange programmes, stricter criteria were implemented. As shown in Table 4, all randomised controlled trials were included, for as explained above, these offer the strongest causal inference to be made. Quasi-experimental studies which undertook a non-equivalent control group design were also included. Consequently, a pre-test measure was collected in these studies from which baseline equivalence could be gauged. Where no pre-test or control group was utilised, studies were excluded. Moreover, where baseline equivalence was not provided or unclear, studies were also excluded. This ensured that the treatment and control group in these studies were matched at baseline in proficiency. Any study designs which contained no control group were also excluded. Studies with no control group have long been lamented in SLA research (e.g., Freed, 1995; Edmonds & Gudmestad, 2018) for as noted by Lafford and Collentine (2006, p. 122) “for those studies lacking an AH [at-home] control group it is difficult to contribute any observable gains (or lack thereof) to the learning condition(s) of the SA experience itself.” Basic time series designs are also excluded, for such a design lacks a suitable control group. As described by Campbell and Stanley (1966), the most basic time-series design introduces one intervention into the data series. While this may be an improvement over the one-group, pre/post-test design, it must still be considered an extremely weak design, and one which holds major threats to internal validity (Kratochwill & Levin, 1978). On the other hand, interrupted control time-series designs are included for they contain control groups and as such one can ascertain whether changes seen in the intervention group are mirrored by that of a control group. Again, baseline equivalence was important. Moreover, ex-post-facto control group designs were excluded due to the lack of a pre-test. These designs occur where participants are measured only on completion of the exchange programme and then compared with a control group who did not go abroad. Moreover, studies which followed a pre-experimental design were also excluded. Examples of such designs include one-group post-test only design (i.e., the treatment group is measured post-study abroad and changes compared to general expectations), and one group pre-test, post-test design. Both designs are susceptible to numerous threats to validity and ultimately cannot answer questions regarding efficacy with any certainty because a control group has not been included. Lastly, observational studies (e.g., cross-sectional designs) were also excluded, again because one cannot determine causality from them. Furthermore, the single time-point serves as both the pre-test and post-test, and it impossible to ascertain whether selection bias existed.

Table 4: Inclusion/Exclusion Study Design Criteria (non-ERASMUS)

Included Designs	Excluded Designs
Randomised Controlled Trials	
Quasi-experimental designs: <ul style="list-style-type: none"> • non-equivalent control group pre-test/post-test Time-series design: <ul style="list-style-type: none"> • interrupted time-series design • control time-series design with equivalent or non-equivalent control group 	Quasi-experimental designs: <ul style="list-style-type: none"> • a non-equivalent control group post-test-only Time-series design: <ul style="list-style-type: none"> • basic time-series designs Ex post facto control group designs Pre-experimental design (pre-post or before and after) Observational design

4.2.4.2 ERASMUS

In order to capture relevant ERASMUS based studies, inclusion criteria were less strict than for non-ERASMUS studies. The rationale for doing so was explored above, and as shown in Table 5, no study design was excluded. As per non-ERASMUS studies, all randomised controlled trials were included as were quasi-experimental designs which utilised a control group. Again, without the control group, measuring the effectiveness of a treatment group is not possible, so this is a crucial component to any included design. Time-series designs were also included, although again, a control group was required. ERASMUS focused studies could be observational or pre-experimental. Both designs were excluded in non-ERASMUS based studies and reflected the less strict criteria employed for ERASMUS based studies.

Table 5: Inclusion/Exclusion Study Design Criteria (ERASMUS)

Included Designs	Excluded Designs
Randomised Controlled Trials	
Quasi-experimental designs: <ul style="list-style-type: none"> • non-equivalent control group pre-test/post-test Time-series design: <ul style="list-style-type: none"> • interrupted time series design • control time-series design with equivalent or non-equivalent control group 	Case study
Pre-experimental design (pre-post or before and after) Observational design	

In brief, the designs included differ in the degree in which causality can be inferred. Those which use randomised controlled trial designs are deemed most methodologically robust, while the

non-equivalent control group design can be considered the most robust within the quasi-experimental group of designs. All other study designs are methodologically weaker, and with it, the ability to infer causality lessened.

4.2.5 Electronic searching

Before conducting the main search, the search strategy was developed by the primary researcher. In brief, each search strategy began with a combination of keywords (e.g., sojourn, study abroad) combined via Boolean Phrasing. Later search strings were a combination of previous search strings in order to narrow down the number of results presented. The search strategies were quality assured by an expert in conducting systematic reviews. Electronic searching took place over one-week, commencing on the 8th of April 2019. The following databases were searched: Web of Science; Article First; ECO; British Education Index; ERIC; PsycInfo; PsycArticles, ProQuest and Scopus. The wide range of databases selected reflects the broad appeal of study abroad research. As a topic, it can stem across many disciplines (e.g., Applied Linguistics, Sociology, Psychology) and as such, it was felt a broad approach to databases selected was required. No time-period was defined, or publication type excluded. This limited potential bias associated with ‘dissemination bias’. An example search strategy can be found in Figure 1, while all search strings can be found in Appendix B.

Figure 1: Web of Science Search Strategy

Search	Terms	Results
S12	S11 AND S2 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	6
S11	S7 AND S10 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	341
S10	S6 AND S4 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	357
S9	S8 AND S6 AND S4 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	0
S8	S5 AND S2 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	503
S7	TS=(<i>*abroad OR sojourn*AND listening OR read* OR writ* OR oral OR spe* OR communi*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	14,511,399
S6	TS=(<i>“language develop*” OR SLA OR “Second Language Acquisition” OR “L2 develop” OR linguistic OR profici*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	146,041
S5	TI=(<i>“language develop*” OR SLA OR “Second Language Acquisition” OR “L2 develop” OR linguistic OR profici*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	40,051
S4	TS=(<i>“study abroad”* OR “year abroad” OR ERASMUS OR “residenc* abroad”</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	6,299
S3	TI=(<i>“study abroad”* OR “year abroad” OR ERASMUS OR “residenc* abroad”</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	3,773
S2	TI=(<i>random* control* trial* OR rct* OR trial* OR review* OR intervent* OR quasi* experimental* OR meta analys*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	1,178,237
S1	TS=(<i>sojourn* OR student* OR “language learner*” OR university*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	1,133,720

Number in bold represents number of final hits imported into Endnote

Once each search strategy was completed and conducted, papers were imported in Endnote X8 for Mac and grouped according to the database name. De-duplication also occurred in Endnote, and once finalised results were imported into specialist software designed for processing studies in a review: EPPI-Reviewer 4 (Thomas, Brunton, & Graziosi, 2010).

A bibliographic search was conducted on the one known meta-analysis (Yang, 2016) to identify any potential papers not found through the electronic searches. Further searches were conducted using the following journals: ‘Frontiers: The Interdisciplinary Journal of Study Abroad’ (<https://frontiersjournal.org>); ‘Study Abroad Research in Second Language Acquisition and International Education’ (<https://benjamins.com/catalog/sar>) and ‘EUROSLA

Yearbook' (<https://www.benjamins.com/catalog/eurosla>). The following volume was also screened: 'Social Interaction, Identity and Language Learning during Residence Abroad' (<http://www.eurosla.org/eurosla-monograph-series-2/social-interaction-identity-and-language-learning-during-residence-abroad/>).

4.2.6 Screening and data extraction process

4.2.6.1 Abstract screening

All titles and abstracts of the located publications were screened against the inclusion/exclusion criteria. All citations were independently double screened to quality assure any decisions. IM screened all, while NB and CT screened half each. If any disagreement could not be resolved by the two reviewers, it was sent to a third reviewer for arbitration.

4.2.6.2 Screening of full papers

Once publications were found, a full screening of texts was undertaken. All citations were screened by IM, while NB and CT screened half of the citations each. As before, in any instances of disagreement could not be resolved, a third reviewer arbitrated.

4.2.6.3 Data extraction

For all included papers, information on programme design, study design, participants, experimental and control conditions, outcome measures, results and study quality were extracted and compiled in a standard format. One paper was extracted by all three reviewers to ascertain inter-rater agreement and ensure the extraction tool served its purpose. IM data extracted from all papers, while CT and NB extracted a sample of two studies each. In instances of any disagreement, a third reviewer was called on to arbitrate. Data were extracted using a data extraction tool, designed for this review, together with being quality appraised. The tool for data extraction included items concerning the nature of the intervention (e.g., accommodation type, sojourner role); description of study design and grouping allocation procedure if relevant; participants and attrition in both groups; outcome measures; results and conclusions.

There was no restriction on the number or type of outcome to be extracted. Outcomes ranged across all linguistic aspects including oral, writing, reading, listening, syntactic, pragmatic and vocabulary learning. Measures had to be standardised instruments, and any self-report data (i.e., perceived change) were not included at the data extraction stage. Studies were quality appraised, particularly in reference to how samples were formed, and how groups were allocated. This was

done in order to establish potential bias, and again in instances of discrepancies, the two reviewers discussed any disagreements.

4.2.6.4 Mapping synthesis

Studies which were carried forward to the third stage of screening were entered into a mapping table (see Appendix D). Its purpose was to collate, describe and catalogue studies which met the inclusion criteria. The mapping table informs readers of the key characteristics of each study, including study design, intervention type, outcome and instruments, and lastly, an overview of the key findings. Studies are listed alphabetically within the wider prism of language project. Many of studies included originate from such projects, and therefore, while they use a subset of participants available to them, the underlying sample available in each study is the same.

4.2.6.5 In-depth review synthesis

Studies which followed only the most robust designs (e.g., RCT, QED) were brought forward to the in-depth review. In doing so, its purpose was to detail studies of causal inference and to account for the counterfactual. For Study 1, the counterfactual was considered non-sojourning language learners. These designs allow for results to be compared to the counterfactual through means of Difference-in Difference (i.e., comparing the difference in pre/post-test scores of the two different groups. Of importance at this stage was confirming baseline equivalence of the outcome measure (i.e., linguistic proficiency). If studies were RCT in design, baseline equivalence was not required for it has been considered superfluous by the CONSORT statement (de Boer, Waterlander, Kuijper, Steenhuis, & Twisk, 2015), for non-equivalence can be considered down to chance. Nonetheless, in any other design (e.g., QED's), baseline equivalence was required. Baseline equivalence was determined by authors providing a non-significant p-value for the outcome measure, although this practice alone may be insufficient to adequately prescribe baseline equivalence. It would have been preferable for authors to report effect sizes for such a statistic allows the reader to determine the magnitude of baseline difference. Moreover, in small sample sizes, p-values alone may present misleading inferences due to having limited statistical power, whereas, in well-powered trials, significant differences may only capture small, often meaningless, differences (Sullivan & Feinn, 2012). However, in several of the studies carried forward to the in-depth review, effect sizes were not provided, nor was it was possible to calculate them from the available raw data, Lastly, it is acknowledged that authors may control for these pre-test differences in the outcome measure (e.g., by using ANCOVA), and were this to be accounted for, more studies could have potentially been carried forward.

Of the included studies, key characteristics were first identified and then tabulated (e.g., study design and study outcome). This was followed by a mapping synthesis, which, like the map found in Appendix D, provided an overview of each of the included studies selected for the in-depth review. Lastly, the evaluation of the methodological quality of each study was undertaken in a two-step process. Firstly, a risk of bias tool was specifically designed for the purposes of this review which explored aspects such as possible sampling bias and grouping assignment. Ultimately, this assessment looked to ascertain whether the conclusions were a fair reflection of the methodology implemented. Secondly, a review of methodological strengths and weaknesses were provided for each study.

Both the structure and findings of this systematic review have been reported in accordance with the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA). PRISMA serves as a tool which outlines an evidence-based set of guidelines and items for reporting systematic reviews and meta-analyses. In doing so, PRISMA aims to help authors improve the reporting of their systematic reviews and ensure that these reviews are transparent and complete in their reporting. In order to achieve this transparency, the tool is made up of two components. The first is a 27-item checklist (see Appendix E) which identifies items which should be reported in the review. The second is a four-phase flow diagram (See Figure 2), the purpose of which, is to show numbers of identified records, excluded articles, and included studies (Liberati et al., 2009). The PRISMA statement is used by roughly 174 journals in the field of health science, and in a sample of 146 journals publishing systematic reviews, the PRISMA statement was referred to by 27% of authors (Tao et al., 2011). It has, nonetheless, seen growing usage in the field of Education, where education policy is beginning to be driven by an evidence-based approach which requires the use of such reviews (Shamseer et al., 2015).

While PRISMA can serve as a suitable guidance tool for many systematic reviews, it provides no guidance for reviews in which a meta-analysis is not required or warranted. As such, the reporting of findings for this review has also been guided by the nine guidelines set out by the *Synthesis without meta-analysis (SWiM) in systematic reviews* (Campbell et al., 2020) reporting tool (see page ...). These guidelines must not be seen as a replacement to PRISMA but designed rather to complement and expand on items concerned with “synthesis of results” (e.g., PRISMA items 14 and 21). Such guidelines are required because, in instances where synthesis methods have not been made transparent, doubts may be cast over the validity of findings. Campbell et al. (2020, p. 1) further note “serious shortcomings” in how narrative syntheses are reported including lack of

methodological description, non-transparent links between study-level data and text synthesis, and inadequate reporting of review limitations. The tool itself was developed in the health care research literature, and to the best of my knowledge, this is the first systematic review in the field of education to implement the SWiM guidelines.

4.3 Results

4.3.1 Searching procedure

4.3.1.1 Initial search

The initial electronic search (w/c 8th April 2018), and the grey literature search, resulted in 2,548 potentially relevant studies being located. Table 6 displays the searched databases, and the number of hits found. Once de-duplicated, 1,533 citations were imported into EPPI Reviewer.

Table 6: Electronic Search Results Prior to Deduplication

Web of Science	First Search		EBSCO				ProQuest	Scopus	Grey Lit
	Article First	Eco	ERIC	British Education Index	PsycArticles	PsycInfo			
357	20	75	603	140	145	285	474	383	66

4.3.1.2 1st stage screening (screening on title and abstract)

The 1,533 imported citations were first screened on title and abstract using the criteria as per Table 3. At this stage, 1,347 (87.8%) were excluded, with 186 citations being brought forward to be screened at stage two. Upon the completion of the 1st stage of screening (title and abstract), the agreement rate for IM and CT and IM and NB was 90%. The remaining 10% of disagreements were resolved through email exchanges between the two reviewers, and where a resolution was not possible, a third reviewer arbitrarily resolved the matter.

4.3.1.3 2nd stage screening (screening on full text)

One hundred eighty-six studies were either located or requested for stage two screening. The full publications of 159 studies were found, with the other 27 papers not accessible (e.g., due to non-permission) within the timeframe. A further three studies were dropped at the 2nd stage of screening due to duplication. Both CT and NB screened half the studies each, with the agreement rate reaching 90% between all reviewers. In order to ensure consistency, CT screened all quasi-experimental design studies.

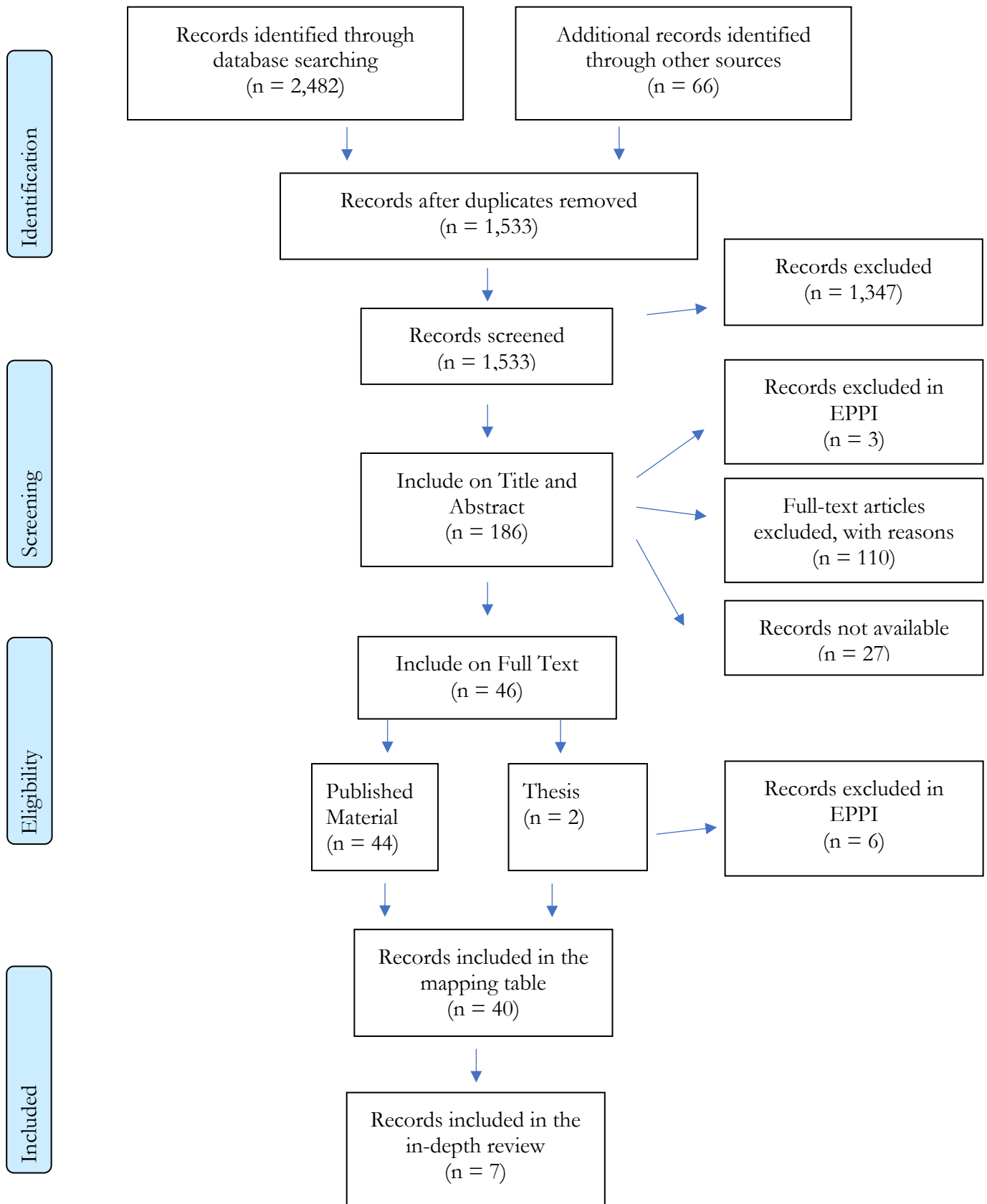
4.3.1.4 *3rd stage screening (data extraction)*

Forty-six studies were included at the final stage of screening; 44 of which were journal articles, and two which were theses. During this stage, six publications were excluded for varying reasons. Four were excluded on intervention as the setting was not conducive to the research question. Within the SLA literature, there are three learning contexts. The first is study abroad and refers to a period living outside the country where formal instruction traditionally takes places. Secondly, there is at-home, which is a semi-intensive intervention. This context can be considered as the traditional control group setting where individuals continue with their normal domestic formal instruction. Lastly, the third context is that of at-home, which is an intensive intervention. This context provides the learner with more opportunities to interact in the L2 than the semi-intensive context. In this context, learners may be discouraged from using the L1, have all extra-curricular activities in the L2 and have access to only L2 materials, including outside the classroom. In sum, the contexts differ in the amount and rate of opportunities available to use the L2. It can be argued whether an ‘at-home intensive’ intervention or an ‘abroad’ intervention offers a more immersive experience, but for the purpose of Study 1, and the thesis as a whole, attention is given to comparing the contexts of ‘abroad’ and ‘at-home, semi-intensive’ interventions. Two studies were excluded on study design, as both were non-ERASMUS studies with one using a cross-sectional design, and another having no control group. Forty studies were included in the map (see Appendix D). All seven included in the in-depth review used a quasi-experimental design. Figure 2 shows the flow of the screening procedure in accordance with PRISMA instruction.

4.3.1.5 *Quality assurance*

Concerning the 40 included studies, two forms of quality assurance of data extraction were undertaken. Firstly, CT quality assured a random selection of three papers in which were in the full mapping table but were not included in the in-depth review. Secondly, full data extraction occurred in the seven studies included in the in-depth review. IM data extracted from all seven studies, while CT and NB extracted from three papers each. Consequently, over 50% of the in-depth review papers were quality assured by two members of the team. Data extraction took place using a tool specifically designed for this review (see Appendix C), and completed tools were emailed to the primary reviewer. Here, the primary reviewer compared the level of data extracted, and a final agreed version was produced reflecting the decisions of both reviewers.

Figure 2: PRISMA Flow-Diagram



From: Moher, D, Liberati A, Tetzlaff, J., Altman, G. (2009). The PRISMA Group. *Preferred Reporting Items for Systematic Reviews and Meta Analyses: The Prisma Statement*. PLoS, Med 6(7)

4.3.2 Study characteristics of all included studies

The key characteristics of the 40 included studies can be found in the tables below. Further information on each study including sample characteristics and key findings can be found in the mapping table in Appendix D. Table 7 lists the included studies according to study design, while Table 8 highlights the range of outcomes the included studies investigated.

Table 7: Frequency of Included Studies according to Study Design

Study Design	No.	Author(s)
Quasi-Experimental Design	7	Hessel, (2016); Hessel & Vanderplank (2018); Li (2014); Llanes, & Muñoz (2013); Jochum (2014); Segalowitz & Freed (2004); Serrano, Llanes, & Tragant (2011)
Individual matched data	18	Avello, (2014); Beattie, Valls-Ferrer, & Pérez-Vidal (2014); Edmonds & Gudmestad (2018); Huensch & Tracy-Ventura (2017); Klapper & Rees (2003); Juan-Garau (2014); Juan-Garau & Pérez-Vidal (2007); Juan-Garau, Salzar-Noguera, & Prieto-Arranz (2014); Mitchell, Tracy-Ventura, & McManus (2017b); Mitchell, Tracy-Ventura, & McManus (2017c); Mitchell & McManus (2015); Mora (2014); Pérez-Vidal & Juan-Garau (2011); Pérez-Vidal & Barquin (2014); Rees & Klapper (2007); Tracy-Ventura (2017); Trenchs-Parera (2009); Valls-Ferrer & Mora (2014)
Pre/Post with no Control Group	11	Avello & Lara (2014); Avello, Mora, & Pérez-Vidal (2012); Barron (2019); Hessel (2017); Ife, Vives Boix, & Meara (2000); Lara, Mora, & Pérez-Vidal (2015); Llanes & Serrano (2011); Llanes, Tragant, & Serrano (2012); Milton & Meara (1995); Serrano, Tragant, & Llanes (2012); Regan (1995)
Cross-Sectional	4	Howard (2005); Howard (2006); Howard (2008); Howard, Lemée, & Regan (2006)

Table 8: Frequency of Included Studies according to Study Outcome

Outcome	No.	Author(s)
Oral	18	Avello (2014); Avello, Mora, & Pérez-Vidal (2012); Avello & Lara (2014); Huensch & Tracy-Ventura (2017); Jochum (2014); Juan-Garau (2014); Juan-Garau & Pérez-Vidal (2007); Lara, Mora, & Pérez-Vidal (2015); Llanes & Serrano (2011); Llanes & Muñoz (2013); Pérez-Vidal & Juan-Garau (2011); Mitchell, Tracy-Ventura, & McManus (2017b); Mitchell, Tracy-Ventura, & McManus (2017c); Serrano, Llanes & Tragant (2012); Serrano, Tragant, & Llanes (2012); Segalowitz & Freed (2004); Trenchs-Parera (2009); Valls-Ferrer & Mora (2014)
Writing	9	Llanes & Serrano (2011); Llanes, & Muñoz (2013); Llanes, Tragant, & Serrano (2012); Pérez-Vidal & Juan-Garau (2011); Pérez-Vidal & Barquin (2014); Mitchell, Tracy-Ventura, & McManus (2017b); Mitchell, Tracy-Ventura, & McManus (2017c); Serrano, Llanes & Tragant (2011); Serrano, Tragant, Llanes (2012);
Grammar	10	Edmonds & Gudmestad (2018); Howard (2005); Howard (2006); Howard (2008); Howard, Lemée, & Regan (2006); Juan-Garau, Salzar-Noguera, & Prieto-Arranz (2014); Klapper & Rees (2003); Mitchell & McManus (2015); Rees & Klapper (2007); Regan (1995)
Listening	2	Beattie, Valls-Ferrer, & Pérez-Vidal (2014); Mora (2014)
Vocabulary	3	Ife, Vives Boix, & Meara (2000); Milton & Meara (1995); Tracy-Ventura (2017);
Global Proficiency	6	Hessel, (2016); Hessel (2017); Hessel & Vanderplank (2018); Klapper & Rees (2003); Li (2014); Rees & Klapper (2007)
Reading	1	Li (2014)
Pragmatics	1	Barron (2019)

N.B: Some studies are counted twice as they explore more than one outcome

Table 9 provides the publication dates of the included studies, with these ranging from 1995 to 2019. Table 10 gives the sample size of studies within ranges from very small to large. Sample sizes ranged between 6 and 145. Many studies comprised fewer than 61 individuals.

Table 9: Frequency of Included Studies according to Publication Date

Date	No.	Author(s)
1995	2	Milton & Meara (1995); Regan (1995)
2000	1	Ife, Vives Boix, & Meara (2000)
2003	1	Klapper & Rees (2003)
2004	1	Segalowitz & Freed (2004)
2005	1	Howard (2005)
2006	2	Howard (2006); Howard, Lemée, & Regan (2006)
2007	2	Juan-Garau & Pérez-Vidal (2007); Rees & Klapper (2007)
2008	1	Howard (2008)
2009	1	Trenchs-Parera (2009)
2011	3	Llanes & Serrano (2011); Pérez-Vidal & Juan-Garau (2011); Serrano, Llanes, & Tragant (2011)
2012	3	Avello, Mora, & Pérez-Vidal (2012); Llanes, Tragant & Serrano (2012); Serrano, Tragant, & Llanes (2012)
2013	1	Llanes & Muñoz (2013)
2014	10	Avello (2014); Avello & Lara (2014); Beattie, Valls-Ferrer, & Pérez-Vidal (2014); Jochum (2014); Juan-Garau (2014); Juan-Garau, Salzar-Noguera, & Prieto-Arranz (2014); Li (2014); Mora (2014); Pérez-Vidal & Barquin (2014); Valls-Ferrer & Mora (2014)
2015	4	Lara, Mora, & Pérez-Vidal (2015); Mitchell, Tracy-Ventura, & McManus (2017b); Mitchell, Tracy-Ventura, & McManus (2017c); Mitchell & McManus (2015)
2016	1	Hessel (2016)
2017	3	Hessel (2017); Huensch & Tracy-Ventura (2017); Tracy-Ventura (2017)
2018	2	Edmonds & Gudmestad (2018); Hessel & Vanderplank (2018)
2019	1	Barron (2019)

Table 10: Frequency of Included Studies according to Sample Size

Sample Size	No.	Author(s)
0-20	11	Avello (2014); Edmonds & Gudmestad (2018); Howard (2005); Howard (2006); Howard (2008); Howard, Lemée, & Regan (2006); Jochum (2014); Juan-Garau & Pérez-Vidal (2007); Serrano, Tragant & Llanes (2012); Regan (1995); Trenchs-Parera (2009);
21-40	12	Avello, Mora, & Pérez-Vidal (2012); Avello & Lara (2014); Barron (2019); Ife, Vives Boix, & Meara (2000); Llanes, Tragant & Serrano (2012); Pérez-Vidal & Juan-Garau (2011); Mitchell, Tracy-Ventura, & McManus (2017b); Mitchell, Tracy-Ventura, & McManus (2017c); Mitchell & McManus (2015); Segalowitz & Freed (2004); Tracy-Ventura (2017); Valls-Ferrer & Mora (2014)
41-60	8	Huensch & Tracy-Ventura (2017); Juan-Garau (2014); Juan-Garau, Salzar-Noguera, & Prieto-Arranz (2014); Klapper & Rees (2003); Lara, Mora, & Pérez-Vidal (2015); Llanes & Serrano (2011); Milton & Meara (1995); Rees & Klapper (2007)
61-80	4	Beattie, Valls-Ferrer, & Pérez-Vidal (2014); Li (2014); Llanes & Muñoz (2013); Mora (2014)
81-100	1	Hessel (2017)
+ 100	4	Hessel (2016); Hessel & Vanderplank (2018); Pérez-Vidal & Barquin (2014); Serrano, Llanes, & Tragant (2011)

Table 11 outlines the target language studied. Learners of English were the most frequently studied, while studies included also investigated French, Spanish, German and Chinese.

Table 11: Frequency of Included Studies according to Target Language

Language	No.	Author(s)
English	22	Avello (2014); Avello, Mora, & Pérez-Vidal (2012); Avello & Lara (2014); Beattie, Valls-Ferrer, & Pérez-Vidal (2014); Hessel (2016); Hessel (2017); Hessel & Vanderplank (2018); Juan-Garau (2014); Juan-Garau & Pérez-Vidal (2007); Juan-Garau, Salzar-Noguera, & Prieto-Arranz (2014); Lara, Mora, & Pérez-Vidal (2015); Llanes & Serrano (2011); Llanes, Tragant & Serrano (2012); Llanes & Muñoz (2013); Milton & Meara (1995); Mora (2014); Pérez-Vidal & Juan-Garau (2011); Pérez-Vidal & Barquin (2014); Serrano, Llanes, & Tragant (2011); Serrano, Tragant, & Llanes (2012); Trenchs-Parera (2009); Valls-Ferrer & Mora (2014)
French	10	Edmonds & Gudmestad (2018); Howard (2005); Howard (2006); Howard (2008); Howard, Lemée, & Regan (2006); Huensch & Tracy-Ventura (2017); Mitchell, Tracy-Ventura, & McManus (2017b); Mitchell & McManus (2015); Regan (1995); Tracy-Ventura (2017);
Spanish	5	Huensch & Tracy-Ventura (2017); Ife, Vives Boix, & Meara (2000); Jochum (2014); Mitchell, Tracy-Ventura, & McManus (2017c); Segalowitz & Freed (2004)
German	3	Barron (2019); Klapper & Rees (2003); Rees & Klapper (2007)
Chinese	1	Li (2014)

N.B: Huensch & Tracy-Ventura (2017) is counted twice as they explore both French and Spanish L2 learners

Table 12 highlights the length of stay of the studies included. Here, most studies either investigated linguistic gains over one semester or one academic year. No study investigated change over a period longer than a year.

Table 12: Frequency of Included Studies according to Length of Stay

Length of Stay	No.	Author(s)
Less than 1 semester (~ 2 months)	3	Li (2014); Llanes & Serrano (2011); Llanes & Muñoz (2013)
1 Semester (~3 months)	23	Avello (2014); Avello, Mora, & Pérez-Vidal (2012); Avello & Lara (2014); Beattie, Valls-Ferrer, & Pérez-Vidal (2014); Hessel (2016); Hessel (2017); Hessel & Vanderplank (2018); Ife, Vives Boix, & Meara (2000); Jochum (2014); Juan-Garau & Pérez-Vidal (2007); Juan-Garau (2014); Juan-Garau, Salzar-Noguera, & Prieto-Arranz (2014); Lara, Mora, & Pérez-Vidal (2015); Llanes & Serrano (2011); Llanes, Tragant & Serrano (2012); Llanes & Muñoz (2013); Mora (2014); Pérez-Vidal & Juan-Garau (2011); Pérez-Vidal & Barquin (2014); Serrano, Llanes, & Tragant (2011); Segalowitz & Freed (2004); Trenchs-Parera (2009); Valls-Ferrer & Mora (2014)
2 Semesters (~ 6 months)	4	Avello & Lara (2014); Ife, Vives Boix, & Meara (2000); Lara, Mora, & Pérez-Vidal (2015); Milton & Meara (1995)
Academic year (~ 9 months)	17	Barron (2019); Edmonds & Gudmestad (2018); Hessel (2016); Hessel & Vanderplank (2018); Howard (2005); Howard (2006); Howard (2008); Howard, Lemée, & Regan (2006); Huensch & Tracy-Ventura (2017); Klapper & Rees (2003); Mitchell, Tracy-Ventura, & McManus (2017b); Mitchell, Tracy-Ventura, & McManus (2017c); Mitchell & McManus (2015); Rees & Klapper (2007); Regan (1995); Serrano, Tragant, & Llanes (2012); Tracy-Ventura (2017)

N.B: Some studies are counted twice as they explore multiple lengths of stay

4.3.2.1 *Synthesis of included studies by intervention*

Of the 40 included studies in the mapping table, 37 were concerned with the ERASMUS programme, while the remaining three were conducted within an American academic setting. Below, studies are divided across intervention type and then once more by target language.

4.3.2.1.1 *ERASMUS: L2 English*

Table 11 lists 22 studies which investigated linguistic change within learners of L2 English. Of these, 18 studies (Avello, 2014; Avello, Mora, & Pérez-Vidal 2012; Avello & Lara 2014; Beattie, Valls-Ferrer, & Pérez-Vidal 2014; Juan-Garau 2014; Juan-Garau & Pérez-Vidal 2007; Juan-Garau, Salzar-Noguera, & Prieto-Arranz 2014; Lara, Mora, & Pérez-Vidal 2015; Llanes & Serrano, 2011; Llanes, Tragant & Serrano, 2012; Llanes & Muñoz, 2013; Mora 2014; Pérez-Vidal & Juan-Garau 2011; Pérez-Vidal & Barquin 2014; Serrano, Llanes, & Tragant, 2011; Serrano, Tragant, & Llanes, 2012; Trenchs-Parera 2009; Valls-Ferrer & Mora, 2014) explored proficiency change in L2 English within an L1 Spanish speaking undergraduate population.

Two of these studies (Llanes & Muñoz, 2013; Serrano et al., 2011) employed a quasi-experimental design examining changes in oral and written ability. Those in the intervention group were abroad for roughly one semester and were compared with a control group who remained in domestic formal instruction. Ten studies followed a longitudinal design, employing a design in which the same individual was followed both during formal domestic instruction and for a period of one semester abroad (individual matched data). In one study (Juan-Garau et al., 2014) the outcome measured was grammatical development, two measured changes in listening ability (Beattie et al., 2014; Mora, 2014), two investigated writing ability (Pérez-Vidal & Juan-Garau, 2011; Pérez-Vidal & Barquin, 2014) and seven (Avello, 2014; Juan-Garau, 2014; Juan-Garau & Pérez-Vidal, 2007; Lara et al., 2015; Pérez-Vidal & Juan-Garau, 2011; Trenchs-Parera, 2009; Valls-Ferrer & Mora, 2014) captured change in oral ability. The remaining six studies undertook a pre/post-test with no control group design. Two (Avello & Lara, 2014; Avello et al., 2012) captured pronunciation change over either three or six months. One (Lara et al., 2015; investigated changes in oral ability, comparing change in a short term (three months) and long term (six months) sojourning group. A further three studies (Llanes & Serrano, 2011; Llanes et al., 2012; Serrano et al., 2012) investigated the domains of complexity, accuracy and fluency in the skills of speaking and writing. For the first two studies, the intervention lasted for one semester, while in the last study, the length of stay lasted one academic year.

Three studies (Hessel, 2016; Hessel, 2017; Hessel & Vanderplank, 2018) investigated L2 English gain in an L1 German-speaking population. In two of these studies (Hessel, 2016; Hessel & Vanderplank, 2018), a quasi-experimental design was employed in order to ascertain the efficacy of sojourning on global proficiency level. Those who went abroad were divided into a short-stay group (one semester) and a long-stay group (one academic year), and then compared with a stay at-home group. The remaining study (Hessel, 2017) undertook a pre/post-test design with no control group comparing two intervention groups, both of whom were abroad for one semester.

Lastly, one study (Milton & Meara, 1995) explored English development in various speakers of L1 European languages. This study employed a pre/post-test design with no control group, determining vocabulary change during a six-month stay at an English-speaking university.

Conversely, 15 studies (Barron, 2019; Edmonds & Gudmestad, 2018; Huensch & Tracy-Ventura, 2017; Howard, 2005, Howard, 2006, Howard, 2008; Howard et al., 2006; Ife et al., 2000; Klapper & Rees, 2003; Mitchell & McManus, 2015; Mitchell et al., 2017b; Mitchell et al., 2017c; Rees & Klapper, 2007; Regan, 1995; Tracy-Ventura, 2017) measured proficiency in a range of L2 European languages within an L1 English undergraduate population.

4.3.2.1.2 ERASMUS: L2 French

Ten studies explored proficiency development in L2 French. Five of these employed a longitudinal design with individuals being matched on their own data both at-home and abroad. Of these, one study investigated vocabulary change (Tracy-Ventura, 2017), two studies (Edmonds & Gudmestad, 2018; Mitchell & McManus, 2015) explored grammatical change, and the remaining two studies (Huensch & Tracy-Ventura, 2017; Mitchell et al., 2017b) explored oral and written proficiency development in the domains of complexity, accuracy and fluency. Five studies (Howard, 2005, Howard, 2006; Howard et al., 2006; Howard, 2008; Regan, 1995) investigated sociolinguistic features of language (e.g., subjunctive, negation and subject-verb agreement). While Regan (1995) employed a pre/post-test with no control group design, all other included studies used a cross-sectional design, comparing learners who had studied abroad for nine months and those who had either remained at-home or not yet begun their sojourn.

4.3.2.1.3 ERASMUS: L2 German

Three included studies investigated change in German (Barron, 2019; Klapper & Rees, 2003; Rees & Klapper, 2007). Two studies (Klapper & Rees, 2003; Rees & Klapper, 2007) employed a

longitudinal, individuals matched data design, following the same individual over four years, both at-home and abroad. In both studies, individuals were measured on grammatical and global proficiency measures with the intervention lasting for one academic year. Barron (2019) undertook a pre/post-test design with no control group, establishing changes in apologetic behaviour (pragmatics). Learners were abroad for one academic year, roughly nine months.

4.3.2.1.4 ERASMUS: L2 Spanish

Three studies (Huensch & Tracy-Ventura, 2017; Ife et al., 2000; Mitchell et al. 2017c) investigated changes in Spanish L2 proficiency in L1 English undergraduates. Ife et al. (2002) undertook a pre/post-test with no control group, capturing change in vocabulary in a group who went abroad for three months, and a group who went abroad for six-months. Both Huensch and Tracy-Ventura (2017) and Mitchell et al. (2017c) employed a longitudinal, individuals matched data design measuring proficiency after three years of formal domestic instruction, and one year spent abroad. While Huensch and Tracy-Ventura (2017) investigated change in oral fluency only, Mitchell et al. (2017c) explored changes in both oral and written proficiency.

4.3.2.1.5 Non-ERASMUS

Three studies (Jochum, 2014; Li, 2014; Segalowitz & Freed, 2004) investigated linguistic gain in sojourners of whom were not partaking in the ERASMUS exchange programme. All three studies used a quasi-experimental design and were as such brought forward to the in-depth review. Both Jochum (2014) and Segalowitz & Freed (2004) investigated oral proficiency gains in an L1 English undergraduate population, of who were studying L2 Spanish abroad for one semester. Li (2014) explored proficiency changes in reading and global proficiency in L1 English learners of Chinese. These learners spent two months in China.

4.3.2.2 Summary of mapping table

The mapping table consists of 40 studies, ranging across several linguistic domains. The purpose of the mapping table was to capture the current state of evidence regarding SLA literature which met the inclusion/exclusion criteria. While the primary question was not exchange programme specific, given the less strict criteria implemented for ERASMUS programmes, the mapping table gives particular focus to the ERASMUS programme (37 out of 40 studies captured focused on a European sample), something which has often been missing in past published narrative reviews (e.g., Llanes, 2011). This also reflects the recency in which authors have investigated the

ERASMUS programme (notwithstanding early research such as Meara, 1994; Milton & Meara, 1995).

The findings of the seven studies carried forward to the in-depth review (see section 4.3.3) will be reviewed in more detail later in this chapter but do nonetheless contribute to the findings provided in this summary.

Improvement in oral fluency has been a consistent finding across the studies, regardless of the language learnt, learners L1 or length of stay, suggesting this is the skill that benefits the most from a period abroad. Indeed, studies of only three months in length (e.g., Juan-Garau & Pérez-Vidal, 2007; Lara, Mora, & Pérez-Vidal, 2015) have demonstrated short-stays abroad are sufficient to result in substantial improvement. Similarly, oral accuracy was shown to develop as a result of a sojourn period, although the extent of this change was often smaller than that of oral fluency. Pronunciation demonstrated less conducive evidence of change (e.g., Avello, Mora, & Pérez-Vidal, 2012; Avello & Lara (2014), possibly because improvement in this measure is measured using subjective means (e.g., using native L2 speakers).

Writing is also a skill which has been well-researched in the identified studies (e.g., Llanes & Serrano, 2011; Mitchell et al., 2017b; 2017c). Overall, such studies have found sojourning to have a positive impact on writing ability both in written fluency and written accuracy. It should, however, be noted, that gains in ability tend to be found in those studies which capture change over a year-long period as opposed to a shorter, three months stay abroad, suggesting oral skills develop at a faster rate than writing skills (e.g., Mitchell et al., 2017b; 2017c; Serrano, Tragant, & Llanes, 2012).

Reading and listening were found to be rather understudied in the included literature. Where researched, however, the results were promising. Beattie, Valls-Ferrer and Pérez-Vidal (2014) found participants to make significantly greater progress in listening ability during a period abroad as opposed to being at home. No European study explored the skill of reading.

While vocabulary was consistently found to improve during a period abroad, in relation to the size of both active and receptive vocabularies (e.g., Ife, Vives-Boix, & Meara, 2000; Tracy-Ventura, 2017), grammatical change has produced more mixed evidence. One potential reason behind this, as demonstrated by the mapping table, is the specificities to which grammatical

ability is measured. For example, while Edmonds and Gudmestad (2018) found target-like rates of gender marking to substantially improve during the period abroad, whereas Howard (2008) demonstrated little improvement in the use of the subjunctive in spoken French. In relation to grammar, it can perhaps best be summed by stating that those forms learnt late by native L2 speakers, continue to allude L2 non-native speakers even after a period abroad (Howard, 2008).

One study, Barron (2019), explored pragmatic change on a study abroad, again demonstrating mixed findings. Exploring the use of apologies in the L2 (German), Barron noted change in all three directions; while some apologies moved towards more native norms, others demonstrated no change, while others moved away from the target-norm. Barron (2019, p. 103) suggests these findings indicate that learner's apologetic routine competence to be "complex" and "non-homogenous". Given that only one study was found to explore pragmatics, the mapping table draws few conclusions regarding the extent to which study abroad improves pragmatic competences.

General proficiency, as often measured by a C-test, has demonstrated significant change during a period abroad. This has been evidenced in both longitudinal studies (e.g., Rees & Klapper, 2007) and in a more traditional group comparison design (e.g., Hessel & Vanderplank, 2016). Given its importance to the thesis, this finding is discussed in more detail in Chapter 6.

In summary, the mapping table has demonstrated evidence to indicate that a sojourn, and more specifically that of the ERASMUS programme, can positively influence several linguistic domains. While these findings may reflect the present consensus of the impact sojourning can have on SLA based on previous narrative reviews (e.g., Borràs & Llanes, 2019; Llanes, 2011), it is hoped the focus of ERASMUS studies, together with the methodological scrutiny and originality applied to these selected studies can offer value to the SLA field.

Given the multitude of study designs presented in the mapping table, the question as to whether these gains can be directly attributed to the sojourn experience remain. For this a further in-depth review is required, highlighting findings from those studies which have used an adequate stay-at-home control group, a study design which has traditionally been used as a barometer in ascertaining the extent to which study abroad contributes to linguistic gain (Grey, 2018).

4.3.2.3 *Narrative synthesis: the in-depth review*

Seven studies (from the 40 above) warranted further investigation and were put forward to the in-depth review. These studies were carried forward because they undertook a pre/post, quasi-experimental design, which as aforementioned, apart from RCTs, provide the most robust evidence that the change witnessed is due to the intervention itself. Any study other than RCT in design must have demonstrated baseline equivalence on the outcome measure. It is possible that more studies would have been carried forward to the in-depth review, had the criteria allowed for studies which controlled for baseline differences within later analysis, although the precise number is not known.

4.3.2.3.1 *Descriptive information of all in-depth review studies*

The four tables below present descriptive information of all the included studies in the in-depth review. As shown in Table 13, all seven utilised a longitudinal, quasi-experimental design. Outcomes were varied across studies (Table 14) with some studies investigating more than one outcome. In total, four investigated oral changes; three investigated writing changes; one investigated reading development; two focused on aspects of grammatical development; two focused on vocabulary, and lastly, three studied overall global proficiency development. No included study investigated listening or pragmatic development during a study abroad period. Formation of the treatment group varied across included studies, as highlighted in Table 15. In two studies, individuals had to go through an application process and were successful in applying for a sojourn programme. This application was competitive and based on several set criteria, including grades, CV and performance in an interview. In a further two studies, individuals received a scholarship to study abroad, although the process involved in gaining this scholarship or how it may have related to the application process above were not known. For the remaining five studies, it was unclear how the treatment group was formed; for example, individuals may have been successful in applying to go abroad or have volunteered to study abroad. Table 16 presents how the control group was formed. Two studies formed their control group from individuals who were unsuccessful in applying for entry onto the sojourn programme, for unknown reasons. In five studies, it was unclear how the control group had been formed. It was unclear if any (unsuccessful) application process was undertaken or whether individuals had decided not to study abroad voluntarily.

Table 13: Study Designs (In-depth Review)

Study design	No.	Author(s)
Quasi-experimental, pre-post	7	Hessel (2016); Hessel & Vanderplank (2018); Jochum (2014); Li (2014); Llanes & Muñoz (2013); Segalowitz & Freed (2004); Serrano, Llanes & Tragant (2011)

Table 14: Study Outcomes (In-depth Review)

Outcome	No.	Author(s)
Oral	4	Llanes & Muñoz (2013); Jochum (2014); Segalowitz & Freed (2004); Serrano, Llanes & Tragant (2011)
Writing	2	Llanes & Muñoz (2013); Serrano, Llanes & Tragant (2011)
Reading	1	Li (2014)
Grammar	2	Llanes & Muñoz (2013); Serrano, Llanes & Tragant (2011)
Vocabulary	2	Llanes & Muñoz (2013); Serrano, Llanes & Tragant (2011)
General Proficiency	3	Hessel (2016); Hessel & Vanderplank (2018); Li (2014)

Table 15: Formation of the Treatment Group (In-depth Review)

Treatment group formation	No.	Author(s)
Successful application (e.g., interview, grades) onto exchange programme	2	Hessel, 2016; Hessel & Vanderplank (2018)
Received a scholarship	2	Llanes & Muñoz (2013); Serrano, Llanes, & Tragant (2011)
Unclear how treatment group was formed	3	Jochum (2014); Li (2014); Segalowitz & Freed (2004)

Table 16: Formation of the Control Group (In-depth Review)

Control group formation	No.	Author(s)
Failed application onto study abroad programme so remained at-home	2	Hessel, 2016; Hessel & Vanderplank (2018)
Unclear how control group was formed	5	Jochum (2014); Li (2004); Llanes & Muñoz (2013); Segalowitz & Freed (2004); Serrano, Llanes, & Tragant (2011)

Table 17 below provides an overview of the seven included studies. Each study is described in relation to its design, intervention type, outcomes, instruments used, participants, and its key findings.

Table 17: Mapping Table (In-depth Review)

Study and Author	Study Design	Intervention	Outcomes and Instrument	Participants	Key Findings
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	Pre-test/post-test, quasi-experimental design. Non-random group assignment. Participation voluntary.	ERASMUS exchange programme. L1 German learners of English studying at an L1 English speaking university. Intervention group consists of those who stay abroad for 3 months and 9 months. Control group consists of domestic-based learners who failed in their application onto ERASMUS.	General L2 proficiency measured via a c-test	143 L2 learners of English split across three groups. Short stay ($n = 45$); Long-stay ($n = 54$); Control ($n = 44$). Mean previous experience of English: 8.69. Average starting proficiency of all groups was B2 (upper intermediate)	Both YA groups experienced significant improvement in overall L2 proficiency across 3 months. AH group made no significant change during this time. Long-stay group maintained proficiency gains – significant difference between T2 and T3. Ah group also made significant gains between T2 and T3 and between-group differences at T3 were not significant
Hessel, G., & Vanderplank, R. (2018). What difference does it make? Examining English proficiency gain as an outcome of participation in ERASMUS study abroad programmes in the UK	Pre-test/post-test, quasi-experimental design. Non-random group assignment. Participation voluntary.	ERASMUS exchange programme. L1 German learners of English studying at an L1 English speaking university. Intervention group consists of those who stay abroad for 3 months and 9 months. Control group consists of domestic-based learners who failed in their application onto ERASMUS.	General proficiency as measured via a C-Test	136 L2 learners of English split across three groups. Short stay ($n = 44$); Long-stay ($n = 52$); Control ($n = 40$) Average previous learning experience = 8.69 of formal domestic instruction. All at B2 (Upper intermediate) at the beginning of the intervention.	Both YA groups experienced significant improvement in overall L2 proficiency across 3 months. AH group made no significant change during this time Long-stay group-maintained proficiency gains – significant difference between T2 and T3. Ah group also made significant gains between T2 and T3 and between-group differences at T3 were not significant

Llanes, A., & Muñoz, C. (2013). Age Effects in a Study Abroad Context: Children and Adults Studying Abroad and at-home	Pre-test/post-test, quasi-experimental design. Group assignment non-random. Participation voluntary.	ERASMUS exchange programme. L1 Spanish learners of English spending 2 or 3 months living in an English-speaking country	Oral and written complexity, accuracy and fluency. Writing: Produce two descriptive essays (one pre-test, one post-test) with learners given 15 minutes to complete task Oral: Picture-elicited narrative task	66 L2 learners of English split across two contexts SA ($n = 46$); AH intensive ($n = 20$) Mean age of all adults = 20.9; average age of onset = 8.42 and all had received over 1,620 hours of formal instruction	SA context more beneficial than the AH context for the improvement in oral skills but no real difference in writing skills
Serrano, Llanes, A., & Tragant, E. (2011). Analysing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs study abroad in Europe	Pre-test/post-test, quasi-experimental design. Group assignment non-random. Participation voluntary.	ERASMUS exchange programme. L1 Spanish learners of English spending time either abroad or at-home Intervention group consists of those who stay abroad for 3 months. Control group consists of two types of domestic learner: a) semi-intensive AH b) intensive AH – groups differ on contact hours with L2	Oral and written complexity, accuracy and fluency Writing: Produce two descriptive essays (one pre; test, one post-test) with learners given 15 minutes to complete task Oral: Picture-elicited narrative task	131 L2 learners of English split across three groups. SA ($n = 25$); AH intensive ($n = 69$); AH semi-intensive ($n = 37$)	No differences between intensive AH and SA groups on writing and oral measures Compared to AH semi-intensive, SA group significantly developed more in written and oral productions in terms of fluency and lexical complexity
Li, L. (2014). Language Proficiency, Reading Development, and Learning Context	Pre/post-test, quasi-experimental design. Group assignment non-random. Participation voluntary.	Non-ERASMUS exchange programme L1 English learners of Chinese Intervention group stays abroad for eight weeks. Control group remains in formal domestic instruction (1 semester)	Chinese language global proficiency test capturing components of listening, grammar, reading, translation and writing Reading comprehension test: 10 multiple choice questions	73 L2 learners of Chinese split across three proficiency groups. SA group ($n = 35$); mean age 20.7. AH group ($n = 38$); mean age 21.4	Proficiency: Beginners showed no change in either context. Sojourners in Intermediate and Advanced groups significantly improved more than AH group Reading: Significant difference in Intermediate group only

Jochum, C. (2014). Measuring the Effects of a Semester Abroad on Students' Oral Proficiency Gains: A Comparison of At-home and Study Abroad	Pre/post-test, quasi-experimental design. Group assignment non-random. Purposive sampling based on a set criterion with students volunteering to partake.	Non-ERASMUS exchange programme L1 English learners of Spanish studying at an L1 Spanish speaking university. Intervention group stays abroad for 3 months. Control group remains in formal domestic instruction	Oral proficiency as measured by the Oral Proficiency Interview (OPI)	18 L2 learners of Spanish. SA group ($n = 9$); AH group ($n = 9$) All had completed 1 – 8 semesters of Spanish study with mean proficiency at Int-low for both groups.	SA group improved their oral proficiency more than those AH and at post-test significant between-group differences were found.
Segalowitz & Freed (2004). Context, contact, and cognition in oral fluency acquisition	Pre/post-test, quasi-experimental design. Group assignment non-random. Purposive sampling based on a set criterion with students volunteering to partake.	Non-ERASMUS exchange programme L1 English learners of Spanish studying at an L1 Spanish speaking university. Intervention group stays abroad for 3 months. Control group remains in formal domestic instruction	Oral proficiency: Oral Proficiency Interview (OPI) Cognitive tasks include attention control and lexical access	40 L2 learners of Spanish. SA group ($n = 22$); mean age 20.68 AH group ($n = 18$); mean age 23.39	SA students made significant gains in five of the eight oral measures – OPI, Turn, Rate, Filler-free, and Fluent-run; students in the AH did not show significant gain in any of these measures. Interaction effect for Turn; Rate; and Filler-free suggesting that the SA changed significantly more than AH group

Table 18: Methodological Quality of Included In-depth Studies

Studies	Grouping Strategy	Sampling bias	Attrition/impact on results	Pre-specified outcomes	Are conclusions fair?
Hessel (2016)	Allocation dependent on learning context – students self-selected into learning context	Comparison group formed of individuals who wanted to be in treatment group but were unsuccessful in application. Prior baseline differences considered minimising sampling bias on results.	Low attrition (5.5% - 11.4%). Possible predictors of attrition not discussed although attrition highest in AH group. Impact to results not discussed – unclear if link between proficiency and attrition – states drop out due to ‘personal reasons’. No other reasoning given.	No	Yes

Hessel & Vanderplank (2018)	Allocation dependent on learning context – students self-selected into learning context	Comparison group formed of individuals who wanted to be in treatment group but were unsuccessful in application. Prior baseline differences considered minimising sampling bias.	Attrition not reported	No	Yes
Llanes & Muñoz (2013)	Allocation dependent on learning context – students self-selected into learning context	Authors used all available ERASMUS students to take part. These students had been successful in gaining a scholarship. Unclear whether the AH group chose to stay at-home or were unsuccessful in gaining a scholarship.	Attrition not reported	No	Yes
Serrano, Llanes & Tragant (2011)	Allocation dependent on learning context – students self-selected into learning context	Authors used all available ERASMUS students to take part. These students had been successful in gaining a scholarship. Unclear whether the AH group chose to stay at-home or were unsuccessful in gaining a scholarship.	Attrition not reported	No	Yes
Li (2014)	Allocation dependent on learning context – students self-selected into learning context	Sample divided according to proficiency bands (e.g., intermediate SA vs intermediate AH) no other pre-existing differences controlled for.	Attrition not reported	No	No
Jochum (2014)	Allocation dependent on learning context – students self-selected into learning context	Presence of sampling bias noted by author, but no measure or analysis undertaken to account for this presence.	Attrition not reported	No	Yes
Segalowitz & Freed (2004)	Allocation dependent on learning context – students self-selected into learning context	Steps were undertaken to minimise sample bias, with participant recruitment purposive, based on a criterion. Stronger inference on the role of learning context in results.	7 individuals dropped from sample although it is not clear why.	No	Yes

4.3.2.4 *The methodological quality of in-depth review studies*

As all studies were quasi-experimental in design, the minimum standard of rigour required to evaluate an intervention, in this case, sojourning, was met. Although quasi-experimental design studies can warrant inferences concerning the efficacy of one condition over another, given that there were no RCTs included in the in-depth review, caution must be given to the strength of arguments and conclusions reached. This is because only RCTs can fully control for confounding variables and offer a level of certainty that change witnessed is indeed causally associated with the learning context (i.e., home or abroad) rather than being due to an unmeasured variable. Table 18 above presents the quality appraisal results undertaken for each of the seven studies. All quality appraisals were based on the results of the quality assessment tool used for the in-depth review studies. All included studies presented conclusions which were a fair reflection of the data analysis undertaken, while no study pre-specified any primary outcome.

Across all the in-depth included studies, there were two methodological limitations. The first, as aforementioned, relates to the lack of RCT designs resulting in group assignment which was non-random or self-selected: individuals self-selecting themselves into either the treatment or control group based on their academic programme or own preference to study abroad. In other words, the researcher had no influence on which groups, individuals entered into. Self-selection brings with it bias, for individuals who choose to participate in the study may be systematically different from those who choose not to participate (Christensen et al., 2014). These students may naturally be more motivated as language learners or have higher linguistic proficiency than the 'average' individual. Researchers must be aware of how self-selection bias can influence results. In the instance of the ERASMUS programme, many learners gain scholarships to partake and consequently can be considered 'unique or different' compared to those who do not get the scholarship. One way to mitigate selection bias is to offer an incentive or reward for participation. Individuals are heterogeneously motivated (Hsieh & Kocielnik, 2016). That is to say; some individuals find reward by partaking in the study (i.e., intrinsic motivation), while others require other incentives such as financial compensation (i.e., extrinsic motivation). By incentivising, it is hoped that a sample is indicative of the wider population, and not comprised of only those who may have a strong interest in the topic under study. The use of an incentive was described in only one study (Hessel, 2016). Secondly, all studies have treated the learning context as distinct and separate. In each, those who study abroad are assumed to be different because, in theory, it is an *immersive* context, which provides ample opportunity for L2 interaction. In truth, it is questionable how immersive the sojourning context and this level

differs quite significantly from one learner to another (e.g., Wilkinson, 2002). Whereas some make great effort to interact in the L2, others spend little time abroad using the target language outside the university or work setting. This is further reflected by the myriad of accommodation options and modes of learning available to sojourners. Rate and type of language contact should be considered as important covariates in understanding the role study abroad plays on language acquisition and to address the influence this plays on L2 acquisition. Only one study (Segalowitz & Freed, 2004) accounted for language contact outside the formal classroom, where it was found that differential gains were not predicted by language contact.

Hessel (2016)

The study looked to ascertain the value of sojourning on one's global proficiency compared to remaining at-home. 143 L1 German of English were divided non-randomly between three groups. Fifty-four individuals studied abroad for one year (group one), 45 studied abroad for one semester (group two), and 44 individuals remained at-home (group three). Assignment to each group was voluntary and based on self-selection, although the control group was made up of 'potentially mobile' students who were unsuccessful in their application to study abroad. At baseline, the three groups were well balanced in terms of key participant background variables, their English language learning history, and baseline proficiency scores. The attrition rate was fairly low, (5.5% for the long-term sojourn group, 6.8% for the short-term sojourn group and 11.4% for the non-sojourn group). How this may have biased the results, if at all, was not discussed. For example, less proficient learners may have been more likely to drop-out. Causal inference of the results was made stronger by controlling for possible covariates of change, with findings suggesting language learners became significantly more proficient than if they had remained at-home, particularly over the first three months. However, given the moderate sample size, there is a possibility of Type II error, given that effect sizes for the regression analysis were moderate to large. Sample size also inhibited the number of predictors that could be controlled for within the regression models, and it is, therefore, unknown if learning context would have remained significant, had all significant covariates of change been added into a single model.

Hessel & Vanderplank (2018)

The study's objective looked to establish the effect of learning context (home vs abroad) on the development of L2 English global proficiency. In this study, 96 sojourners were recruited, 52 of whom were studying abroad for two terms or more and 44 who were studying in the UK for one semester. The control group was made up of 40 students who were unsuccessful in their

ERASMUS application and as such, remained at-home. Individuals volunteered to partake in the study, and group assignment was based on sojourn programme (i.e., the authors had no control over which groups each individual was assigned too). Attrition rates were not given. Potential covariates were analysed and then controlled for in the final models, which looked to establish the unique effect learning context played on L2 proficiency development. As these variables were controlled for, it can be said with more certainty the effect of the learning context on proficiency development. As with Hessel (2016), the moderate sample size means only moderate to large effects could be established, meaning the probability of a Type II error is higher.

Llanes & Muñoz (2013)

The objective of this study was to evaluate the effect of learning context (and age) on L2 development in relation to oral and written ability. The authors non-randomly assigned 139 individuals, 66 of whom were adults and 73 were children. Within the adult sample, 20 stayed at-home, and 46 studied abroad. Participation in the study was voluntary, and no attrition was reported. Effect sizes were calculated using Ferguson's (2009) criteria and where given were small to moderate. Given this, the chances of a Type II error are increased due to the relatively small sample. While not necessarily required for the research question, it would have been helpful for statistically significant MANCOVA analysis to be followed up with posthoc tests. This would ascertain how the four groups were different from each other. For example, within the paper, the authors have provided a list of who made the largest improvement but did provide any statistical detail, so it was unclear whether differences between the groups were significantly different. Posthoc analysis follow up is recommended after MANCOVA analysis (Pituch & Stevens, 2016). The findings suggested that overall, study abroad benefited learners more than at-home, particularly in the domain of oral fluency. However, the instrument used, particularly for the writing measure consisted of a timed 15-minute composition of at least seven lines long. It is questionable how sensitive this measure is to ability and whether a one-timed, 15-minute composition is adequate to truly measure one's writing ability. Put differently, is this composition of seven (or more) lines enough to get a fair representation of one's ability?

Serrano, Llanes, & Tragant (2011)

The aim of the study was to examine how learning context influences the development of L2 English written and oral competencies by comparing changes in a study abroad group ($n = 25$) and a group of non-sojourners ($n = 37$). The study also explored changes in a group of 69 L2 learners within an intensive domestic setting, which provided greater L2 immersion than the

semi-intensive context. For the purpose of this review, however, these findings were dropped from the analysis. Justification for the small sojourner sample was given as it was stated that the number of applications onto the ERASMUS programme was lower and less competitive for that academic year. Participation in each group was voluntary, although it is unclear how individuals were assigned to each group. No attrition was described. Thirty-seven individuals performed the post-test writing measure, while only 12 completed the oral task. This difference was explained because of 'practical reasons', and the 12 individuals were chosen at random. The findings suggested that study abroad facilitated L2 oral and written acquisition compared to remaining at-home (semi-intensive). However, because of the two-control group design, those in the treatment group conducted the post-test twice, once 15 days after pre-test, and once again two months later. As such, a task repetition effect may be present, which favoured the performance of the sojourner group as the instruments were very similar.

Li (2014)

The objective of this study was to explore development in reading and global proficiency ability among learners of Chinese in two learning contexts. 73 L1 English learners of Chinese were non-randomly assigned to two groups dependent on learning context and then divided across three proficiency levels. Sample sizes were relatively balanced across the learning contexts and proficiency levels, although group numbers were small, ranging from nine to 15 students. All individuals came from the same university, which the author described as *prestigious*.

Consequently, the generalisability of findings is limited and may not be representative of learners studying at a less academic institution. The findings suggested that sojourning had a positive impact on both reading and general proficiency with the study abroad group scoring significantly higher on both at pre-test ($p < .05$). However, given the small sample sizes, there is a high chance of a Type II error. No effect sizes were given. Moreover, it is questionable whether the measures, particularly that of reading, capture the construct it was designed for. Reading was measured through 10 multiple-choice, comprehension questions, with a top mark of 10. It is questionable whether this measure is sensitive enough to capture actual reading ability and given the high scores achieved by groups at pre-test (range between 9.12 and 9.51), it is highly likely the measure is limited by a ceiling effect. Put differently; both groups were shown to have high reading ability prior to the intervention occurring or that the measure was too easy, decreasing the likelihood the testing instrument has accurately measured the construct.

Jochum (2014)

The aim of this study was to determine whether a study abroad experience contributed to the development of oral proficiency in a group of 18 L2 Spanish adult learners. Participants were all enrolled at the same university, divided into two groups of nine according to the learning context. Justification for the small sample was given in that the population available to the researcher was not large. Given that all students came from the same university and the small sample size, generalisability is an issue and is not a fair representation of the wider population. Purposeful criterion sampling was used to select participants so that the two groups were matched on several variables, including learners having no previous experience abroad, and all learners majoring or minoring in Spanish. The findings suggested that sojourning had a beneficial impact on global oral proficiency with a one-way ANOVA indicating a post-test significant difference ($p = <.05$) between the two learning contexts. Nonetheless, while the study looked to determine the effects of one semester abroad (12 weeks), individuals did not complete the instruments at the same time. The pre-test was completed within six weeks of arrival abroad, while the post-test was completed within six weeks of arrival home. Whether the result of the test is a true reflection of change while abroad is questionable as scores achieved may not be reflective of their ability at pre-and post-experience abroad. Moreover, participants were told of their results within two weeks of taking the pre-test. This may further bias the results as individuals who performed badly may be less motivated in the post-test. Lastly, the small sample size increases the possibility of a Type II error.

Segalowitz & Freed (2004)

The objective of the study was to investigate the role of the learning context in the successful acquisition of L2 Spanish oral skills. Participants were 47 native speakers of English studying Spanish for one semester in one of two contexts, either at-home or abroad. Purposeful criterion sampling was used to select participants with a final sample consisting of 40 individuals, 18 of whom remained at-home, and 22 who studied abroad. As such, seven participants were dropped from the recruited sample, although it is unclear why these seven individuals did not meet the criteria. Due to the small sample size, there is a high possibility of a Type II error. While abroad, individuals participated in a minimum of three language courses per week, compared to those at-home who participated in just one. It can therefore not be discounted that gains seen in oral fluency were not resultant of the learning context per-se, but simply that those abroad experienced more in-class L2 contact hours. Nonetheless, the results of the study suggested that the study abroad context can lead to greater oral gains than remaining at-home.

Table 19: Synthesis Without Meta-analysis (SWiM) Items

SWiM reporting item	Item description	Page in manuscript where item is reported	Other*
Methods			
1 Grouping studies for synthesis	1a) Provide a description of, and rationale for, the groups used in the synthesis (e.g., groupings of populations, interventions, outcomes, study design)	80-82	
	1b) Detail and provide rationale for any changes made subsequent to the protocol in the groups used in the synthesis	N/A	
2 Describe the standardised metric and transformation methods used	Describe the standardised metric for each outcome. Explain why the metric(s) was chosen and describe any methods used to transform the intervention effects, as reported in the study, to the standardised metric, citing any methodological guidance consulted	N/A	
3 Describe the synthesis methods	Describe and justify the methods used to synthesise the effects for each outcome when it was not possible to undertake a meta-analysis of effect estimates	N/A	
4 Criteria used to prioritise results for summary and synthesis	Where applicable, provide the criteria used, with supporting justification, to select the particular studies, or a particular study, for the main synthesis or to draw conclusions from the synthesis (e.g., based on study design, risk of bias assessments, directness in relation to the review question)	72	
5 Investigation of heterogeneity in reported effects	State the method(s) used to examine heterogeneity in reported effects when it was not possible to undertake a meta-analysis of effect estimates and its extensions to investigate heterogeneity	N/A	
6 Certainty of evidence	Describe the methods used to assess the certainty of the synthesis findings	75	
7 Data presentation methods	Describe the graphical and tabular methods used to present the effects (e.g., tables, forest plots, harvest plots)	87-90	
	Specify key study characteristics (e.g., study design, risk of bias) used to order the studies, in the text and any tables or graphs, clearly referencing the studies included		
Results			
8 Reporting results	For each comparison and outcome, provide a description of the synthesised findings and the certainty of the findings. Describe the result in language that is consistent with the question the synthesis addresses, and indicate which studies contribute to the synthesis	97-104	
Discussion			
9 Limitations of the synthesis	Report the limitations of the synthesis methods used and/or the groupings used in the synthesis and how these affect the conclusions that can be drawn in relation to the original review question	106-107	

4.3.3 Results of the in-depth review

As aforementioned, the seven studies were selected due to their study design, from which stronger causal inference can be made. The certainty of evidence is based on the risk of bias assessment tool (Table 18) and the in-depth methodological review for each study as per above.

Findings have been organised as per linguistic outcome. This grouping strategy was chosen because of the homogeneous characteristics of the interventions found in each of the seven studies. Linguistic outcomes are presented in a hierarchical fashion reflective of the SLA literature. The domains of oral, writing, and reading are first specified and serve as the foundational skills of language learning. Under oral and writing, this skill domain is divided across 'fluency' and 'accuracy'. The skills of grammar and vocabulary are then discussed, followed by general proficiency, which can be considered an overarching proficiency measure. All included studies investigated the role study abroad (intervention) had on the development or ability of a particular linguistic skill (outcome). The findings are presented using the guidelines provided by SWiM, as shown above in Table 19.

4.3.3.1 Oral

Oral proficiency

Only one study (Jochum, 2014) investigated the role sojourning had on one's global proficiency development in a non-European population. A computerised version of the Oral Proficiency Interview was administered to 18 individuals, who took a pre-test and post-test, roughly one semester apart. At pre-test, no significant difference was found between the two learning contexts. At post-test however, a one-way ANOVA determined that mean proficiency in the abroad group was statistically significantly higher than that of the at-home group ($F(1,16) = 7200, p = .016$). Moreover, paired sample t-tests showed a significant average improvement in the abroad group ($t(8) = -3.59, p = .007$), but inconclusive evidence of improvement in the at-home group ($t(8) = -2.29, p = .051$). Jochum (2014) concluded, therefore, that the study abroad was beneficial to the development of oral proficiency, more so than remaining at-home. Given the small sample size, which is acknowledged by the author, the study is underpowered, and the potential for Type II error increased. Given the methodological shortcomings of this study, it can be described as a weak study, and one cannot say with any great certainty, that sojourning had a positive impact on global oral proficiency.

Oral fluency

Three studies (Llanes & Muñoz, 2013; Segalowitz & Freed, 2004; Serrano, Llanes, & Tragant, 2011) explored the domain of oral fluency, demonstrating how communicatively competent the individual is when speaking. All three considered change over a one-semester period.

Segalowitz and Freed (2004) explored the role studying abroad played on one's oral fluency change through the means of eight variables in 40 (18 at-home; 22 abroad) L2 learners of Spanish over one semester. A series of paired sample t-tests were conducted, where it was found that sojourners had significantly improved in four of the eight variables; namely Turn, Rate, Filler-Free and Fluent-run all at $p = <.05$ level. Moreover, significant interaction effects were identified, indicating that sojourners improved significantly more than non-sojourners in Turn ($p = .007$, $\eta^2 = .17$), Rate ($p = <.001$, $\eta^2 = .42$) and Filler-free ($p = .028$, $\eta^2 = .12$). This suggested that a semester abroad was sufficient time to develop and that on return home, sojourners spoke more words within each speech turn, spoke more words per minute and used fewer filler words when speaking than those who remained at-home. The effect sizes are relatively large, particularly for rate, although again, the certainty regarding the evidence is questionable due to the small sample. It is likely that, although large effect sizes were found, the study remains underpowered and there is a possibility of a Type II error.

Both Llanes and Muñoz (2013) and Serrano, Llanes and Tragant (2011) explored one construct of oral fluency, namely that of syllables per minute (similar to 'Rate' above). Llanes and Muñoz (2013) first explored paired sample t-tests finding adult sojourners to become significantly more orally fluent at post-test ($p <.001$), producing on average 23 syllables per minute. A MANCOVA was further conducted finding sojourning adults had higher mean gains than non-sojourning adults, although it is not reported whether this difference was statistically significant, and no effect size given. Serrano et al. (2011) undertook a MANCOVA analysis finding sojourners to use significantly more syllables per minute than those at-home at post-test ($p = .034$, $\eta^2 = .13$).

Oral accuracy

Two studies (Llanes & Muñoz, 2013; Serrano et al., 2011) investigated the impact of sojourning as an outcome on oral accuracy, with both studies measuring accuracy by calculating the number of errors per t-unit. Llanes and Muñoz (2013) conducted paired samples t-tests finding a non-significant change in oral accuracy in learners of both contexts. MANCOVA analysis showed at-home adults to make stronger gains than adult sojourners in oral accuracy; however, no statistical

description is given, and as such it is unclear whether this difference is significant, or just how large the difference in improvement between the two groups was. The MANCOVA analysis conducted by Serrano et al. (2011) found a non-significant difference in gains achieved by adult learners in both contexts, concluding gains in oral accuracy to be comparable between the two learning contexts. Interestingly, those abroad regressed in their oral accuracy score at post-test compared to pre-test, while those at-home showed a slight, albeit non-significant improvement. No *p*-values were given for this change.

In synthesis of the above findings, despite the analysis of the included studies being underpowered, there was relatively uniform improvement in the oral fluency of sojourners after one semester abroad. Oral accuracy, on the other hand, showed little change, with the included studies indicating that a) formal domestic instruction may be the superior learning context in facilitating acquisition and b) a semester abroad is an insufficient length of time to stimulate substantial improvement. Nonetheless, the robustness in evidence concerning these statements is weakened by the limited number of included studies, the limited sample size, and the use of only one variable to measure oral accuracy. Furthermore, concerning the development of global oral proficiency and phonological acquisition, the certainty of evidence is weak and robust conclusions cannot be drawn from the available evidence.

4.3.3.2 *Writing*

Writing fluency

Two studies (Llanes & Muñoz, 2013; Serrano et al., 2011) evaluated the impact of sojourning on writing fluency over one semester as a result of learning context. In both studies, fluency was measured via words per t-unit. Undertaking paired-sample t-test analysis, Llanes and Muñoz (2013) found that individuals both at-home and abroad to improve non-significantly over a three-month period abroad in writing fluency. A follow-up MANCOVA showed at-home adults to make the largest gains in writing fluency, but due to no posthoc follow-up, it is unknown whether this gain was significantly larger than adult sojourners, as the paper also explored changes in children. Serrano et al., (2011) undertook a MANCOVA analysis, finding sojourners to score significantly higher at post-test in written fluency than non-sojourners ($F(1,60) = 4.12$, $p = .047$, $\eta^2 = .069$), which is a medium partial eta squared effect size. At post-test, those abroad achieved a mean score of 12.15, while those at-home scored 10.02. The standard deviations of these scores in both learning contexts were larger at post-test, suggesting greater individual differences at post-test.

Writing accuracy

Two studies (Llanes & Muñoz, 2013; Serrano et al., 2011) investigated the impact study abroad has on changes in writing accuracy over a three-month period within an L1 Spanish speaking sample. Llanes and Muñoz (2013) found neither group to show a significant gain over time in paired sample t-test analysis. MANCOVA analysis also demonstrated no significant main effect for learning context or age; neither did it find a significant main interaction, indicating learning context has little added advantage in the development of writing accuracy. Similarly, Serrano et al. (2011) found a non-significant interaction effect in their MANCOVA, indicating learner performance in the two contexts was not significantly different from each other.

When synthesised, the two studies support the argument that sojourning has little impact on writing accuracy, and given the homogeneity between the two studies, the certainty of evidence is relatively strong for writing accuracy. Writing fluency, on the other hand, produced mixed results, and therefore, there are no clear conclusions that can be drawn from this review concerning the development of writing fluency. Given the small sample sizes of both studies, the possibilities of a Type II error are increased, and one must be mindful that both fluency and accuracy were measured by a single construct on a small dataset (compositions of seven + lines).

4.3.3.3 Reading

Only one study (Li, 2014) investigated how study abroad can influence reading development. Reading ability was measured by ten multiple-choice comprehension questions. Individuals were split across learning context (home vs abroad) and proficiency type (beginner, intermediate, and advanced). At pre-test, there were no significant differences in reading proficiency between learning contexts and proficiency banding was matched across learning context (i.e., intermediate learners were of similar proficiency both home and abroad). At post-test, Li undertook a 2x3 ANOVA, where post-test scores served as the dependent variable, while proficiency banding and learning context served as independent variables. A significant effect was found, indicating that the scores of those abroad were significantly higher than the at-home groups $F(3, 69) = 11.137$, $p = .003$. Further simple main effect analysis was conducted on each proficiency band to identify the pattern of differences. Here, it was found that at beginner level, sojourners did not outperform those at-home $F(1, 24) = 64.238$, $p = 42.09$; nor did advanced learners $F(1, 20) = 12.804$, $p = 2.103$. For intermediate learners, on the other hand, the effect was significant $F(1, 27) = 41.382$, $p = <.001$, indicating that those abroad came home with significantly higher

reading ability than those at-home. At post-test, intermediate learners scored a mean reading score of 9.32 compared to a score for those at-home of 8.58.

In synthesis, the paper found sojourning to benefit the reading ability of intermediate learners most, but given that only one study focused on reading, and the methodological weaknesses inherent in this study, the robustness of evidence must be considered weak. The primary issue in this study is the sensitivity of the measure used and the subsequent ceiling effects which were found at pre-test.

4.3.3.4 *Grammar*

Two studies (Llanes & Muñoz, 2013; Serrano et al., 2011) also explored syntactic complexity as measured by clauses per t-unit and refers to the range and sophistication of syntactic structure employed in speech or writing.

Llanes and Muñoz (2013) found a non-significant improvement in syntactic complexity across both domains (oral and writing) and learning context, as measured by a paired sample t-test. This indicated that neither group scored significantly higher at post-test than pre-test in either domain. Further MANCOVA analysis indicated a non-significant interaction effect between age and learning context for oral syntactic complexity. For writing syntactic complexity, on the other hand, a significant interaction effect was found, where at-home adults made the largest gains. While this interaction effect is significant, no further analysis was conducted, and as such, it is unclear whether a significant difference in scores adult sojourners and non-sojourners existed. Serrano et al. (2011) found no evidence in their MANCOVA to suggest that sojourners outperformed non-sojourners in syntactic complexity as the difference in post-test scores between both contexts were non-significant in oral and written domains.

The overall evidence indicates that a period abroad of three to 12 months has only a small impact on the development of grammatical aspects, particularly those which are acquired late within a native speaker population (e.g., the subjunctive). Given the reliance on cross-sectional designs and small sample sizes, robust evidence is limited, the strength of conclusions weakened.

4.3.3.5 Vocabulary (Lexical Richness)

No study has specifically explored vocabulary change as the primary outcome. Nonetheless, two studies (Llanes & Muñoz, 2013; Serrano et al., 2011) explored lexical complexity (vocabulary sophistication) in both oral and written output using Giraud's Index of Lexical Richness.

Regarding lexical complexity in writing, MANCOVA analysis conducted by Serrano et al. (2011) indicated a significant effect ($F(1, 60) = 19.62, p = .001$). This indicated that at post-test, sojourners had significantly more sophisticated vocabularies than those who remained at-home in both their writing. Paired sample t-tests analysis by Llanes and Muñoz (2013) found at-home adults to improve significantly in written lexical complexity from pre to post-test, $t(1, 17) = -3.38, p = .004$, although a non-significant difference was found in the sojourner sample. A non-significant MANCOVA analysis indicated that for written lexical complexity, those who lived abroad did not gain significantly more in written lexical complexity compared to those who remained at-home.

Concerning oral complexity, Serrano et al. (2011) found a significant interaction between learning context and age existed for oral lexical complexity ($F(1, 35) = 4.32, p = .046$). The paper identifies that study abroad adults had the highest gains, followed by at-home adults, child sojourners and lastly at-home children. Unfortunately, due to non-inclusion of follow up posthoc analysis, it is unclear whether the gains seen in the adult sample were significantly different (i.e., were gains in the adult sojourn sample significantly higher than at-home adult sample, or the at-home/abroad child sample). The interaction effect size was also relatively small (.05), according to Ferguson's (2005) criteria. Llanes and Muñoz (2013) conducted paired sample t-tests revealing non-significant differences over time for oral lexical complexity in both learning contexts.

In sum, the findings suggested sojourning to have a positive effect on vocabulary development, particularly within the oral domain. Given again, the homogeneity of the two studies, and both finding that oral lexical complexity grew as a result of sojourning, it can be said with more certainty that sojourning does have an effect. More mixed results were found concerning writing lexical complexity, and it cannot be said with any certainty that sojourning benefits this outcome any more than remaining at-at-home.

4.3.3.6 General Proficiency

Three studies (Hessel, 2016; Hessel & Vanderplank, 2018; Li, 2014) evaluated global proficiency development as a result of a sojourning experience. General proficiency concerns itself with a more macro perspective of linguistic ability as opposed to a particular domain. In two studies (Hessel, 2016; Hessel & Vanderplank, 2018), global proficiency was measured via a C-Test, while Li (2014) used a specifically designed holistic measure, testing varying language skills such as listening and translation. The results of Hessel and Vanderplank (2018) have originated from Hessel (2016) study, with the same sample and results reported. Consequently, to avoid repetition, I will report, in detail, the results of Hessel (2016), although this also holds true for Hessel and Vanderplank (2018) also. Comparing gains made in a short-term sojourn group, a long-term sojourn group and an at-home group, Hessel (2016) first undertook independent sample t-tests in order to measure change. Her results indicated that during the first period three months abroad, both the short-term group ($p = <.001$) and long-term group ($p = <.01$) made greater mean average gains than the at-home group. Analysis was then conducted on controlling for a series of variables (e.g., learner attitude, motivation and learning characteristics) in order to ascertain whether gains witnessed could be attributed to the learning context after controlling for other such factors. Regression analysis showed that even after controlling for these factors, learning context remained a significant predictor of language change. Further analysis was conducted on the following six months (January – June) where it was found that both the abroad group ($p = <.001$) and the at-home group ($p = <.01$) showed a significant change in proficiency, although, for the abroad group, this gain had plateaued over the six months. Moreover, learning context was no longer a significant predictor of language change after controlling for potential covariates. The findings indicated a beneficial effect of sojourning on global proficiency but suggested short-term sojourning may hold more advantage than long-term sojourning.

Using a general proficiency test, comprised of various linguistic components (e.g., listening, translation), Li (2014) split her sample between both learning context and across proficiency bands. A 2x3 ANOVA was performed on post-test scores with sojourners showing significantly higher scores on average compared to non-sojourners $F(3, 69) = 8.781, p = .004$. Simple main effect analysis showed a non-significant difference for beginners $F(1, 24) = 2.342, p = .101$, but scores for both the intermediate $F(1, 27) = 31.634, p = <.001$ and advanced groups $F(1, 20) = 12.804, p = .003$ were significantly higher in the sojourn sample.

As Hessel (2016) took a number of steps to control for possible confounding variables and given her large sample size in comparison to other studies, Hessel (2016) and subsequently Hessel and Vanderplank (2018) must be considered to hold the most methodological rigour, and with it, these findings are considered with stronger certainty. Li (2014) also showed evidence that sojourning can benefit global proficiency, again showing a short-term period (8-weeks) is a sufficient length of time compared to when remaining at-home. The study had adequate sample size and an instrument which used a number of linguistic aspects to find an overall score as compared to only one measure. Both studies showed gain across multiple languages also (English and Chinese) suggesting generalisability across languages.

4.4 Discussion

There has, to the best of my knowledge, been no previous narrative synthesis of studies investigating the value of sojourning on linguistic development, compared to remaining at-home.

The systematic review has identified 40 studies which met the inclusion criteria, seven of which were carried forward to the in-depth review. Given the low number of included studies, together with the limited nature of this evidence base, inferences concerning the value of study abroad on linguistic development is severely weakened. Nonetheless, two observations are warranted from the limited evidence concerning type and rate of language acquisition abroad.

4.4.1 Linguistic outcome under study

Teachers, administrators and students tend to over-estimate the advantageous nature of studying abroad on language development (Badstübner and Ecke, 2009). This is perhaps not surprising given the accumulation of anecdotal evidence over the years and the continued assumption that language can only be learnt when it is 'lived' in (Hessel, 2017). From the strength of evidence of studies included in this review, the strongest claims for improvement centre on oral fluency and global proficiency. In each of the studies exploring oral fluency, a significant change was witnessed. This supports the notion that sojourners are afforded exposure to, and opportunities to interact in, the target language, and become more communicatively competent as a result (e.g., Pinar, 2018). Moreover, these gains were witnessed in both L2 English and L2 Spanish, suggesting gains may be generalisable across languages. The development in global proficiency is perhaps less conducive to clear explanation but implies that learners have automatized their L2 and require less cognitive load when using the L2. According to the Skill Acquisition Theory, closely linked to the Adaptive Control of Thought model (Anderson, 1982), the acquisition of

any skill goes through a three-stage process. First declarative (new) knowledge is introduced to the learner, second, the knowledge becomes procedural through practice, and lastly, the skill becomes automated. Studying abroad can minimise the time it takes for one to transition from procedural to automated knowledge, whereas remaining at-home can extend it. This further supports the notion, that individuals abroad, receive more opportunity and practice to use the L2, hence the faster development.

The in-depth review found the sojourning context to be less conducive to the development of writing, reading and grammatical skills. Such findings may reflect the limited practice of these skills in a sojourning context as evidenced by the literature (e.g., Dewey, Belnap, Hillstrom, 2013; Isabelli-García, 2006; Mitchell et al., 2017a) In such studies, sojourners have tended to shy away from interacting with the L2 in media forms such as TV, radio and books, instead focusing on oral communication. In a classroom, however, greater emphasis is placed on the accuracies of language, with much effort and attention played to writing, grammar and comprehension. This may also serve as the reasoning behind why in the domain of accuracy, both in speaking and writing, learners tended to develop more in the at-home group. In the classroom, users often receive immediate feedback from the teacher, whereas abroad, evidence suggests that native speakers tend to be slow to correct or offer no correction at all (DeKeyser, 2007). Returning to the Skill Acquisition Theory, this can result in fossilised errors while abroad, which can take tremendous effort to correct when back in the classroom. Lastly, sojourners may be influenced by accent and dialect depending on the region visited, and as such return home with non-standard norms (DeKeyser, 2007) which could influence grades during final year of study.

4.4.2 Length of stay

All seven studies (Hessel, 2016; Hessel & Vanderplank, 2018; Jochum, 2014; Li, 2014; Llanes & Muñoz, 2013; Segalowitz & Freed, 2004; Serrano et al., 2011) measured individuals abroad for one semester, while two (Hessel, 2016; Hessel & Vanderplank, 2018) also measured individuals who had been abroad for the academic year. Overall, there appears evidence that three months is a sufficient length of time to stimulate significant positive language change. Hessel (2016), for example, considered the most robust methodological design, found improvement to somewhat plateau after three months. Moreover, this finding is in line with the meta-analysis conducted by Yang (2016), which found a smaller effect size for long-term study abroad programmes, compared to short-term programmes. There does however appear an important nuance to this point. The seven included in-depth review studies suggested that while three months is sufficient

time for fluency to develop, it appears insufficient for the development of accuracy. This may reflect the purpose of language use abroad, with users requiring to be communicatively competent rather than ‘accurate’.

4.4.3 Policy and practice

We believe the two discussion points above could have important implications for both policy and practice going forward. From a policy perspective, a move towards more short-term study abroad programmes should be promoted. Research (e.g., British Council, 2005, 2017; UUKI, 2019) have consistently found a barrier to participation, particularly in a British setting has been the length of stay. Sojourning can add a further year to a learner’s undergraduate study and create undue financial pressure compared to remaining at-home. With the political environment in the UK, new opportunities may arise to implement different lengths of study abroad programme, better fitting the needs of the current climate of students. For language students, one semester abroad may be enough to promote significant change.

The review has shown and offers practitioners a more nuanced overview of the linguistic domains which benefits most from a period abroad. Suggesting to pre-sojourners that the study abroad stimulate language development can, for some appear somewhat ambiguous and provide unrealistic expectations of just how much change one can expect to witness. If these expectations are not matched, sojourners can often return home disappointed and disillusioned with the L2 (Jackson, 2018; Wilkinson, 1998). This review offers, to some extent practitioners around the globe, to provide more realistic expectations to their language sojourners and just the extent to which to expect change. In doing so, sojourner may leave with more realistic expectations and come back more motivated to continue learning the language.

4.4.4 Strengths and limitations

The strength of this review is that it has tackled a broad and diverse literature field and systematically synthesised the literature. The review has included both European and non-European samples, and when discussing results looked for homogeneity in sample characteristics, outcomes, and the instrument used. Nonetheless, a weakness of this review is that no RCT was found or included. As such, the causal inferences drawn from the study’s conclusions are limited, and it is not possible to be confident that the changes witnessed are due entirely to the learning context. Secondly, given the number of sojourners each year, no included study can be considered large-scale. Within the in-depth review, the largest sample size was 143

(Hessel, 2016). Consequently, the conclusions offer limited generalisability to the broader sojourn population. Moreover, the samples within the in-depth review were mostly proficient learners who already had a high standard of linguistic ability prior to the sojourn experience. The Threshold Hypothesis (Cummins, 1976; DeKeyser, 2007) would dictate that in order to develop while abroad, learners would have already met a certain proficiency level. The results are therefore not generalisable to beginner learners, or even those who are lower intermediate, who may not be able to take advantage of all the linguistic opportunities afforded to them to develop.

4.4.5 Future directions

This review has provided a narrative synthesis of results of studies to account for 'how' language changes as a result of the year abroad. By taking a comparative approach, studies included in the in-depth review have provided measurement of how language changes in one context in relation another. Future systematic reviews should be interested in understanding *why* language changes going beyond learning context and investigating the role of individual differences. For example, previous research (e.g., Dörnyei & Ryan, 2015) has shown numerous individual differences (e.g., learner motivation, learning strategies) to impact the rate of change and account for why one learner may develop more than another. A review to synthesise such literature would provide a useful insight into how practitioners can promote language development during a period abroad.

4.4.6 Conclusions and recommendations

In conclusion, this systematic review has found tentative evidence to suggest sojourning has a positive impact on language development. Causal inferences concerning the role of learning context are limited due to the lack of RCTs and strong quasi-experiments with random or selected allocation to study at-home or study abroad, but the review has shown limited evidence that language change is highly compartmentalised and dependent on the length of stay. There is a pressing need for practitioners to take an evidence-based approach when providing pre-sojourners with realistic expectations of the sojourn experience regarding language change. The review has identified possible policy and practice implications and provided suggestions for future reviews.

- The results have included studies with both European and non-European populations, reflecting the new diversity in research samples emerging in the past decade.
- Given the lack of RCTs for an efficacy-orientated research question, quasi-experimental designs have been the most rigorous designs undertaken in the field.

- The review has utilised guidelines as presented by both PRISMA and SWiM. While PRISMA is increasing in usage in the field of Education, this review is the first to the best of my knowledge to use the SWiM guidelines.
- Undertaking a sojourn experience can enhance linguistic ability, particularly in the skills of general proficiency and oral fluency.
- Length of stay appears to influence differing linguistic constructs. Short stays (~ three months) facilitate skill fluency, but not skill accuracy. Longer stays (~ nine months) aid skill fluency and skill accuracy.
- Short stays also appear to benefit general proficiency, with longer stays providing little additive benefit.
- Findings are generalisable to a particular population: those who are 18+ and are of upper intermediate to lower advance proficiency.
- Further rigorous research is required, particularly in the areas of reading and listening.
- Future reviews should explore further possible predictors (e.g., personality, language attitudes, self-efficacy) on accounting for individual differences in language gain.

This chapter has presented Study 1 of the thesis which has been a systematic review, with the purpose of capturing the value of a study abroad in relation to language proficiency. Chapter 5 will detail Study 2, serving as the longitudinal, empirical component of the thesis.

Chapter 5: Study Two – Longitudinal Study

5.1 Introduction

Study Two serves as the empirical component of this thesis and uses a predominantly quantitative approach to reach its conclusions. Longitudinal in design, the study lasted roughly 16 months, beginning in July 2018 and ending in October 2019. The study's sample comprises of 180 undergraduates, from a range of British universities, 110 of whom went on a year abroad, and 70 of whom remained at-home in their 3rd year of study at their domestic institution.

5.1.1 Background context of the study

Within the UK context, most learners begin their academic studies aged 18, with the typical undergraduate course lasting three years. This can be extended to four if the learner completes a year abroad or undertakes an integrated Masters. While the first year offers a chance to assimilate to the university culture with little academic pressure, a student's academic results in their 2nd and 3rd years of academic study go towards their degree classification. The studied discipline at university will be reflective of the subject path chosen at both school and college, and as such learners will likely have five or more years learning experience in the particular discipline.

For language learners, universities offer two strands of learning pathways. Firstly, learners can enter university with an academic qualification in a particular language (e.g., A-Levels). In my own experience as a language learner, all language instruction will be conducted through the L2, covering both linguistic and cultural topics of the chosen language. Learners can expect roughly five hours of contact in the L2 language per week. On the other hand, universities offer ab-initio courses, where no previous experience of the L2 is expected. These learners will likely receive more contact hours, and teaching will be more directed towards the linguistic aspects of the language. Regardless of the pathway chosen, all learners are expected to be at B2 level, as rated by the Common European Framework of Reference at the end of their second year of study. It is recognised, however, that different institutions employ their own regulations and the information provided above may not be representative of every institution in the UK. It nonetheless provides a rough guideline for the reader.

Given the importance placed on evaluating the learning contexts of study abroad and at-home in this study, it is important to outline from the outset how these two learning contexts may differ.

Year abroad

The year abroad can be considered a sandwich year and serves as a natural hiatus from domestic academic study. For many students, the year abroad takes place in year three, typically lasting for roughly nine months, beginning in September and ending in June. The length of stay varies considerably, often ranging between seven and 13 months. As explored earlier, the year abroad can be described as a *British* phenomenon (Alred & Byram, 2002), for sojourns in other European countries typically last for one semester (Sanz, 2014), while even shorter programmes are the norm in the USA (Kinging, 2015). Participation on a year abroad can be either compulsory or voluntary dependent on degree studied. For language learners in the UK, the experience is compulsory, and learners live in the country of their target language. If a learner is learning two or more languages, then they will split their year abroad accordingly, although this is again dependent on the institution. For courses which do not contain a language component, the year abroad is typically voluntary, and students can choose whether they study abroad or remain at-home in their 3rd year of study. For those who undertake a year abroad, two pathways are available. Learners can either participate through ERASMUS or in an overseas exchange programme. The difference between the two is that while the former is European based, the latter allows for non-European international travel. Overseas exchange programmes are typically more competitive, and funding less readily available.

While abroad, students have a number of roles available to them. Learners can either study at a university, teach English as a language assistant, or work privately. In some instances, students can switch roles during the year, for example, both studying and working. Likewise, sojourners have several living options available, ranging from private accommodation to homestays. Particularly with regards to language students, the options chosen can greatly impact the number of linguistic opportunities afforded, while for non-language students, the options chosen, tend to reflect easy access to English (e.g., choosing modules taught through the medium of English).

At-home

For those who remain at-home, their 3rd year of study represents their final year of academic study at undergraduate level. This can be considered the year with the greatest academic pressure, for scores achieved in coursework and exams hold the highest weighting. While such pressures may confound potential personality change (i.e., increased anxiety), a rationale for selecting 3rd year at-home students is provided in section 5.2.4.1. Learners also typically complete a dissertation during this year, preparation for which usually begins in November. Exams are

held in January and May. Learners usually stay in private accommodation outside the remit of the university and live with friends who in many cases were also housemates in the previous year. As a result, non-sojourners typically have firm social networks already in place and in many instances, learners can quickly adapt to returning to academic study after the summer break. Put differently, the transition from 2nd to 3rd year is relatively inconsequential for many.

5.1.2 Aims and objectives

From an applied perspective, Study 2 looked to evaluate the year abroad as an intervention in fostering positive personal growth and facilitating positive well-being. Moreover, the study looked to establish possible reasoning (i.e., factors) behind why some sojourners may demonstrate stronger trait-like tendencies over time or experience greater positive well-being while abroad than others. Consequently, Study 2 looked to provide both a descriptive and explanatory account of personality growth and well-being during a year abroad compared to remaining at-home. Furthermore, the study looked to ascertain whether sojourning aided the acquisition of a second language. Due to the lack of a comparative group, the results here are only indicative, and no causal inference can be warranted. Moreover, the study tested whether personality served as an individual difference in accounting for differences in change scores found, adding to the sparse literature on the subject matter.

From a theoretical perspective, the study aimed to inform the field of personality psychology, taking a dynamic approach to personality, one which reflects the current thinking of the field. Personality was captured at three hierarchical levels (i.e., broad trait, narrow trait and states) to capture the dynamic interplay of personality constructs. Moreover, at the state level, specifically, the study aimed to add the burgeoning literature on understanding whether variability at the within-person level is systematic and predictive of outcomes measured.

In sum, the study has looked to take a novel approach in assessing whether the goals and purpose of a year abroad are realistically achieved in those who undertake them. Taking a cross-disciplinary approach, the study served to provide original findings in a number of research areas, which, to date, are relatively unexplored.

5.1.3 Rationale

The continued push to widen mobility participation is reflected in the year-on-year increase in exchange programme uptake. Whether undertaken through ERASMUS+ or other overseas

exchange programmes, the experience is repeatedly assumed by many (e.g., King & Ruiz-Gelices, 2003; Jackson, 2008; Sigalas, 2010) to develop personal competencies. These competencies are essential in adapting and thriving in a globalised world and becoming a ‘global’ citizen. The European Commission (2015, p.15) describe the beneficial nature of sojourning as follows:

“Student mobility contributes to individuals’ personal development and equips them with a wide range of competencies and skills that are increasingly valued by employers. Students improve not only their foreign languages skills and develop greater intercultural understanding, but they also become much more able to quickly adapt to changes and new situations, solve problems, work in teams, think critically and communicate more efficiently.”

While these beliefs are well versed, there is a continued need to test these claims through empirical research (Byram, 2008). Many of these beliefs have stemmed from retrospective testimonies of former sojourners rather than more objective, longitudinal means. Both these channels of research should be assimilated to understand the full impact of sojourning on individuals. Firstly, it is vitally important that individuals depart with realistic expectations for the year abroad. If the impact is hyperbolised, then this can ultimately lead to learner disappointment and frustration, compromising the very objectives such programmes look to achieve (DeKeyser, 2014; Jackson, 2018; Wilkinson, 1998). Secondly, given the administrative resources in running and maintaining these programmes, we have a responsibility as researchers, to ensure that these programmes offer value for money and ultimately achieve what they set out to achieve. Without evaluative studies, such insights cannot be gained, and new policies may ultimately fail.

Given the breadth and depth of personality research, personality development on a year abroad is currently under-researched and provides an avenue for new research and exploration. Firstly, past personality research (e.g., Lüdtkte et al., 2011; Niehoff et al., 2017; Zimmermann & Neyer, 2013) have shown precedents in investigating higher-order traits (e.g., the five-factor model) over three or fewer timepoints. Here, samples, have predominantly been specific to German sojourners, while the only British-based study (Tracy-Ventura et al., 2016, p. 122) explored personality change in a treatment group only. The authors note the following:

“Additionally, it would be worthwhile to replicate this study with the addition of a control group that does not go abroad and a comparison group who goes abroad but to a country where the same language is spoken.”

This cited research above has tended to view within person-change as ‘noise’, or random error, and have as such based their conclusions on between-group change. This is in contrast to the anecdotal literature, which explores personality development through the prism of the individual. Here, research is concerned with understanding how individuals retrospectively perceive change while abroad, often focusing on lower-order traits such as anxiety (e.g., Tracy-Ventura et al., 2016), curiosity (e.g., Huhn et al., 2016) and resilience (e.g., Michl, Pegg, & Krachen, 2019). These two approaches should, however, complement one another, investigating change at both the between and within-person level, which this study aims to achieve.

Secondly, young adulthood has received increasing attention within the lifespan/life-event literature due to the turbulent nature of personality development during this life period (e.g., Bleidorn, 2012; Specht, 2017). The role sojourning plays in this is not yet fully realised, although it has been argued that sojourning does serve as a valid life event (e.g., Zimmermann & Neyer, 2013). Little is known regarding the maturation process abroad and how its development may differ from that demonstrated in a domestic context.

Lastly, recent developments in personality theory and methodology enable us to push our understanding of personality change forward. This study takes advantage of combining both the trait and social cognitive perspectives in order to provide a *descriptive* and *explanatory* account of personality change. The study also makes use of experience sampling methodology, which to the best of knowledge, is the first to use this technique in a study abroad context.

In sum, this Study 2 aims to offer original insights, through novel approaches, into the role sojourning plays in personal development in comparison to remaining at a domestic institution.

From a linguistic perspective, learning context has long been recognised as a crucial factor in the successful acquisition of a second language (DeKeyser, 2007; Llanes, 2011). Today, this area of interest is well-researched, yet as aforementioned in Chapter 4, results of this literature have proven inconsistent. Moreover, as much of the available literature originates from an American context, comparison with European learners is made difficult. While recent language acquisition projects (see Mitchell et al., 2017a; Sanz, 2014) have tried to readdress this imbalance, the limited research on European language learners is perhaps surprising given the importance the European Commission places on multilingualism and outward mobility.

The field of SLA has also paid a keen interest in understanding the role of individual differences in language learning, under the notion of the ‘Good Language Learner’ (Rubin, 1975).

Traditionally, personality has received little of this attention, overshadowed by variables such as learning motivation, aptitude and strategies. Dörnyei and Ryan (2015) further note that the SLA field has been unable to keep pace with the advances seen in the personality literature such as new methodologies and theories. Nonetheless, more recent publications (e.g., Marijuan and Sanz, 2018) have called for a more considerable investigation in the role personality plays in accounting for individual differences in language learning. Moreover, conceptualising personality in the SLA literature has proven troublesome. Ehrman (1996), for example, argued that scholars view personality as attributable to *how* a language is learnt rather than *how well*. This is evidenced by Ożańska-Ponikwia and Dewaele (2012) among others, who found openness to be a predictor of language usage, from which they deduced openness to predict language change. There is a need to go a step further than these assumptions and objectively measure both language change and personality together to ensure conclusions formed originate from an evidence-based approach.

5.2 Methodology

5.2.1 Sampling and participants

Data were analysed from 180 British domiciled students, who were either living abroad or at-home during the academic year 2018/19. All students were completing their third year of undergraduate study, having all completed the past two years at their domestic institution.

5.2.1.1 Sampling strategy

The first step was to identify which universities would be eligible to partake in the study. It was decided that universities must be situated in the UK and offer students a year-long study abroad programme. As such, the target population were 3rd year British domiciled undergraduate students, who were either undertaking the academic year abroad (treatment) or continuing to study at their domestic institution (control). Universities which met these criteria were then listed and entry grades (A-Level) for their BA Modern Languages degree programme identified. Based on entry grade requirements, universities were split across three levels. The grade equivalent for each grouping was as follows: high = A*, A; mid = B; low = C, D. Universities were split in this way in order to gather a more generalisable sample, for if focusing on one group or university alone, the findings may be biased and not indicative of the student population at large.

Given the researcher’s own links to the universities of Durham and Swansea, the sampling strategy to select both these universities can be considered as convenient. Based on the grouping strategy employed above, Durham fell into ‘high’, while Swansea fell into ‘mid’. A further three universities were selected into ‘high’ and ‘mid’, while a further four were selected into the ‘low’ group. These ten remaining universities were randomly selected by the researcher. That is to say; there was no practical or theoretical reasoning why one university was chosen over another. These universities were selected using a simple random procedure by a computer randomly choosing universities from a prescribed list.

In total, twelve universities were selected, split across the three groups. In Group A were Durham, Birmingham, Bristol and Exeter. In group B were Newcastle, Reading, Sheffield and Swansea. In group C were Chester, Nottingham Trent, Bangor and Hull. Table 20 displays the selected universities across each group, together with the number of individuals who partook in the study. At the selected universities, relevant departments and points of contact were listed.

Table 20: Overview of Universities Selected

Group	University	Sojourners	Non-Sojourners
A	Durham	47	36
	Birmingham	12	20
	Bristol	26	5
	Exeter	0	4
B	Newcastle	7	1
	Reading	6	1
	Swansea	1	1
	Sheffield	4	0
C	Bangor	4	2
	Nottingham Trent	3	0

5.2.1.2 Recruitment

Once universities were selected, recruitment strategies were employed to find participants.

Contact was first made to the respective departments, via email, once ethical clearance was given in May 2018. The email provided an overview of the study’s research aims together with the ethics certificate. If a department agreed, a follow-up email was sent, containing a pre-written message advertising the study. To maximise outreach to undergraduates, it was asked if study information could be shared via email and on university message boards. To minimise the administrative burden, participants were asked to show interest directly to the researcher.

There was a keen interest shown by the Modern Language Departments in many universities selected. All but Exeter University, who declined to participate and Swansea University, who did not respond, agreed to participate and advertise the study. Hull and Chester were later dropped due to administrative reasons¹. There was, however, less interest shown by non-language departments. Here only the English department at Birmingham, the Psychology department at Bristol, and the Geography department at Durham agreed to advertise the study.

All twelve study abroad offices (group A-C) were also targeted in the recruitment process. These offices process outgoing students (i.e., students preparing for their year abroad) and had access to the bulk email addresses of all relevant students. If no response had been given by the 14th of July 2018, they were no longer contacted. Only Swansea, Bangor and Nottingham Trent declined or gave no response, while all others responded positively.

As the response rate from the universities themselves was reasonably low, it was decided that an expanded recruitment drive would be necessary, which was targeted at advertising on Facebook and Twitter. Here, the study was advertised on discipline-specific and non-discipline specific pages. This ensured the study had maximal outreach to all potential participants, of who were either preparing for international or domestic study. This facilitated a rise in response rate between June and September 2018.

Interest was signalled by the completion of a Google Form, which triggered an automatic email to be sent to the interested party. This email contained a link to complete the first questionnaire and had a 'further information' sheet attached. Individuals were asked to read the information sheet, and if happy, were asked to click the link. This questionnaire began with a consent form, which if accepted, asked participants to complete their demographic information. If the individual did not consent, the questionnaire ended, and no data were collected.

Focus group participants

Focus group sampling took place once the main study had been concluded. Both recruitment and sample composition can be described as *convenience*, as the focus groups were only open to returning students of one institution and who had taken part in the study. Participation was

¹ There was a long delay in processing the administrative requirements needed to pass ethics (e.g., the study had to go through the ethics board of the selected university).

advertised via email only, with 26 individuals showing interest. Seventeen (65%) later gave consent, but three participants were dropped as they could not make any of the sessions.

In light of the strategies employed above, the sample consists of both sojourners and non-sojourners from a range of universities across the country. The techniques employed were targeted at individuals who were in their 3rd year of study and from a range of disciplines. Nonetheless, the sample size is skewed towards Group A universities suggesting that the findings may hold less generalisability to non-red brick university, while convenience sampling also holds that the findings produced are only generalisable to that particular sample (Bornstein, Jager, & Putnick, 2013). Moreover, the sample consisted of a group of volunteers, which bring into question self-selection effects which require further analysis.

5.2.1.3 Response rate

Once the call for study interest was closed, 343 individuals had shown interest in the study. Fifty-seven of these individuals did not consent or failed to respond, leading to a consent rate of 83.3% (286 participants). Twenty-one individuals were dropped in the screening process. Of these, 17 were 2nd year students, 3 were 4th year students, and one had already graduated. During the study, a further 24 formally left the study by emailing the researcher their wish to stop participating. These individuals were removed immediately from all mailing lists and their data securely deleted. Fifty-nine participants were dropped from the full analysis because they had too few data points, averaging two from a possible 18. Twenty-seven of these individuals were sojourners, while the remaining 32 were non-sojourners. Lastly, two participants were dropped because they completed one semester abroad, as they did not return after Christmas for undisclosed reasons. The data analysis is, therefore based on 180 participants. Table 21 details the percentage of response rate per questionnaire and timepoint for the 180 participants.

Table 21: Overview of Instrument Response Rate

Trait Personality											
T1			T2			T3			T4		
178			177			176			163		
99%			98%			95%			91%		
Facet											
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	
155	172	170	177	178	167	177	171	162	166	161	
94%	96%	94%	98%	99%	93%	98%	95%	90%	93%	90%	
State Personality											
November				February				May			
1986				1382				1352			
40%				27%				26%			
Language											
French			German			Italian			Spanish		
T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
34	31	27	19	16	13	11	9	9	35	30	31
94%	86%	77%	100%	84%	68%	100%	82%	82%	95%	81%	84%

5.2.1.4 Sample characteristics

Out of the 180 individuals who remained throughout the study, 110 were sojourners (intervention group), and 70 were non-sojourners (control group). Table 22 indicates that individuals were reasonably comparable across the two groups on demographic characteristics, although there was a gender imbalance in both groups. 16.3% of the sojourner sample were males, while 15.8% of the non-sojourner sample was male, and these figures are in line with the published UK sojourn research (e.g., Mitchell et al., 2017a; UUKI, 2019).

Opening the study up to 2nd year undergraduates was considered. After all, the at-home environment for 2nd year undergraduates may be more reflective of the sojourn environment. For example, in each environment, academic pressure is relatively low. Moreover, opening the study up would have increased the sample size. However, it was decided that the project would only be open to 3rd year undergraduates, so that demographically, the two groups would be as similar as possible. This is important because personality is intrinsically linked with demographic characteristics such as age (e.g., Bleidorn, 2015) and there was a concern that by having potentially younger, less mature, students in the comparison group, the two groups would have been substantially different on the personality measures at baseline. Given the study design implemented, it was essential that pre-test differences could be minimised as much as possible.

Table 22: Overview of Sample Demographic Characteristics

Variable		Sojourners	Non-Sojourners
Setting	Abroad	110	-
	Home	-	70
Gender	Male	18	11
	Female	92	59
Age	-	20.25	20.30
Pre. Experience Abroad	Yes	29	22
	No	81	48
Discipline	Arts and Culture	4	2
	Business	6	7
	English Literature and Language	1	8
	Geography, Sociology and Politics	7	11
	History, Classics and Archaeology	11	12
	Law	6	7
	Medical Science	3	12
	Engineering	4	1
	Maths, Statistics and Physics	2	3
	Natural and Environmental Science	3	7
	Modern Languages	63	-

Within the study abroad group, 63 were language learners. To meet their degree requirement, these learners were required to spend their third year abroad in an L2 speaking country. The languages learnt in the sample are as follows: French ($n = 33$); Spanish ($n = 34$); German ($n = 18$); Italian ($n = 11$); Russian ($n = 2$); Portuguese ($n = 2$) and Japanese ($n = 2$). These individuals were either Single Honour ($n = 29$) or Joint Honour ($n = 34$) language learners. This dichotomy is important for it has implications on how the study abroad is experienced. All Joint Honour students divided their year abroad into two countries, reflecting their learnt languages. In comparison, Single Honour learners either remained in the same country the entire year or lived in two countries which natively spoke the same language (e.g., Argentina and Costa Rica).

Host countries

Table 23 lists all host countries of sojourners in this study. Approximately 70% of sojourners lived in European countries during the year, while the remaining 30% lived outside of Europe. France, Germany, Italy and Spain were the most frequented countries.

Table 23: Overview of Host Countries

1st half (August – January)		2 nd half (February – July)	
Country	Number	Country	Number
Argentina	1	Argentina	3
Australia	3	Australia	3
Austria	1	Austria	5
Belgium	3	Belgium	2
Brazil	-	Brazil	1
Canada	4	Canada	3
Chile	2	Chile	3
China	1	China	1
Colombia	2	Colombia	1
Costa Rica	1	Costa Rica	-
Czech Republic	2	Czech Republic	2
Denmark	2	Denmark	2
France	23	France	19
Germany	13	Germany	8
Greece	1	Greece	1
Hong Kong	1	Hong Kong	1
Iceland	1	Iceland	1
Ireland	1	Ireland	1
Italy	10	Italy	8
Japan	5	Japan	5
Mexico	1	Mexico	-
Morocco	1	Morocco	-
Netherlands	1	Netherlands	-
Portugal	-	Portugal	2
Peru	2	Peru	2
Russia	2	Russia	1
Singapore	2	Singapore	2
Spain	16	Spain	23
Sudan	-	Sudan	1
Sweden	1	Sweden	2
Switzerland	1	Switzerland	1
United States	6	United States	6
% of learners in Europe	71	% of learners in Europe	71
% of learners not in Europe	29	% of learners not in Europe	29

Sojourner departure and arrival home dates

The year abroad programme within the UK context is highly flexible and versatile. While students are expected to remain abroad for a set period for the experience to be classed as a ‘year abroad’, lengths of stay can vary widely. Within the sojourner sample, lengths of stay varied between 7 and 14 months ($m = 10$ months). Individuals departed for their year abroad at

different timepoints, and also returned home at varying times. These are displayed in Tables 24 and 25, with the majority departing in September and returning home in June.

Table 24: Departure Month (Start of the Year Abroad)

Month	Number of individuals	%	Cumulative %
June	2	1.8	1.8
July	12	10.9	12.7
August	33	30.0	42.7
September	48	43.6	86.4
October	15	13.6	100

Table 25: Returning Month (End of the Year Abroad)

Month	Number of individuals	%	Cumulative %
May	15	13.6	13.6
June	53	48.2	61.8
July	22	20.0	81.8
August	13	11.8	93.6
September	7	6.4	100

Table 26 outlines the sojourner role (i.e., studying, working, and teaching) and also the accommodation types of participants. The output indicates that most individuals undertook study during their year abroad, and many lived in student or private accommodation with either native speakers of English or in multilingual dormitories.

Table 26: Sojourner Roles and Accommodation

Variable	Type	Number – 1 st half of YA	Number – 2 nd half of YA
Sojourner Role	Studying	70	63
	Working	20	22
	Teaching	20	25
Accommodation	Student accommodation	41	44
	Private shared accommodation	46	42
	Living alone	12	15
	A homestay	11	9

Note: 15 individuals (16.5%) changed sojourner role, while 44 individuals (40%) changed accommodation type

Focus groups

Participants from one participating institution were invited to attend a focus group discussion. In total, 14 students were able to attend one of the three focus groups. Table 27 shows the characteristics of learners who partook in the focus groups. These individuals were studying a range of disciplines. Roughly half were language learners for whom the year abroad was

compulsory, while the other half volunteered to study abroad. The gender distribution was representative of the study's wider sample as nine were female (71%), and four were male (29%). The length of stay ranged between 7 and 13 months.

Table 27: Overview of Focus Group Participants

Participant*	Gender	Age	Discipline	Length of Stay (in months)	Compulsory /Voluntary
Howard	M	20	Business	9	Voluntary
Ryan	M	23	Geography, Sociology and Politics	10	Voluntary
Michael	M	20	History, Classics and Archaeology	9	Voluntary
John	M	20	Maths, Statistics and Physics	10	Voluntary
Carolyn	F	21	Modern Languages	12	Compulsory
Mary	F	20	Modern Languages	9	Compulsory
Teresa	F	20	Modern Languages	7	Compulsory
Chloe	F	20	Modern Languages	10	Compulsory
Nikki	F	20	History, Classics and Archaeology	9	Voluntary
Laura	F	19	Modern Languages	13	Compulsory
Ella	F	20	Modern Languages	11	Compulsory
Susan	F	19	History, Classics and Archaeology	9	Voluntary
Bea	F	28	Modern Languages	11	Compulsory
Helen	F	20	Modern Languages	11	Compulsory

* Pseudonyms have replaced real names

5.2.1.5 *Minimising missing data*

One of the primary methodological limitations of any longitudinal design is missing data (Carpenter & Plewis, 2011). Such data can cause a decrease in power and precision, together with introducing possible bias, as the sample is not truly representative of the target population (Pampaka, Hutcheson, & Williams, 2016). Various techniques were employed to retain individuals and avoid missing data. Firstly, participants were incentivised to complete each questionnaire round. The first incentive was a personality profile booklet which provided a graphical representation of personality and linguistic change across the year. Participants who had completed the monthly questionnaire were also automatically entered into a prize draw for that particular month. The choice of incentives was driven by BERA guidelines (2011) which discourage the use of money as a reward, together with Harari et al. (2017) findings, that a prize reward proved effective in maintaining participation. Secondly, emails were periodically distributed, informing participants of the study's progress and acting as a general check-up with

participants. Lastly, I created an online forum which participants could choose to sign up to. There was no obligation, and it was designed as a way for all participants to meet each other online and discuss their own issues and experiences of the year abroad. Response and engagement were relatively low, with around 30 individuals showing interest over the year. Missing response data were further minimised using an alert system. This alert was implemented in each questionnaire and informed the user that a missing response was present. The user could then return to the missing question or continue onto the next page.

5.2.2 Materials and data collection

All materials (see Appendices F-K) were distributed throughout the academic year to capture quantitative data concerning personality and linguistic factors. All material was designed and developed based on the literature explored in chapters two and three, together with the relevant methodological literature. All questionnaires were developed, stored and disseminated using software provided by SoSci Survey. This software had several advantages. Firstly, all material automatically adapted depending on the device screen size to give the best user interface experience. Secondly, the software embedded automation tools which provided easy questionnaire dissemination. Thirdly, data could be directly imported into .sav and .csv file formats minimising the need to copy and paste data. Lastly, SoSci Survey has their servers based in Munich and as such adhere to all EU privacy laws and to German data protection laws which prohibits access to or the passing on of third-party data.

5.2.2.1 Broad personality trait questionnaire

Trait personality was captured using the 44-item Big-Five Inventory (henceforth BFI). This self-report questionnaire comprises of short, easy-to-understand, phrases which assess behaviours aligned to each one of the Big-Five domains. For example, “Is curious about many different things”, serves as an openness item, while “Is relaxed, handles stress well”, is a reversed scored item from the neuroticism sub-scale. Respondents were asked to rate themselves on how each item reflects their own behaviours on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The questionnaire has been used with an undergraduate population and has shown high internal consistency and retest reliability (e.g., Rammstedt & John, 2007; Tkach & Lyubomirsky, 2006). From a sojourning perspective, an alternative to the BFI is the Multicultural Personality Questionnaire (Van der Zee & van Oudenhoeven, 2000) which assesses the five traits believed to impact sojourner’s cultural adaptability. The outcomes captured are as follows; Cultural Empathy, Open-mindedness, Social Initiative, Emotional Stability, and Flexibility and have been

correlated with the traditional five-factor model as used in Study 2 (Leone et al., 2005). As noted by van der Zee, Van Oudenhoven, Ponterotto and Fietzer (2013, p. 118), these five traits have “demonstrated incremental validity over broad personality measures such as the Big Five in predicting criteria such as students’ international orientation.” The MPQ has been utilised in sojourning-orientated studies, including Leong (2007) and Tracy-Ventura et al. (2016). However, it was decided that the MPQ would not be used in this study as the measure takes a more intercultural approach to the sojourning experience. Rather than focusing on the moderating effect personality can have on intercultural development, Study 2 aims to gain insights on the impact sojourning has on an individual’s own personality development within the sphere of personality theory and life-event studies.

5.2.2.2 *Narrow trait questionnaire*

A short form of the State-Trait Anxiety Inventory (Spielberger, 1983) developed by Marteau and Bekker (1992) was used to assess anxiety. This short-form version comprises of six statements; three of which are reversed scored on a Likert scale of between 1 (*almost never*) to 5 (*almost always*). An example of an item is as follows; “I am worried”. The short form has been tested on an undergraduate sample of who were not clinically assessed for anxiety disorders. Internal reliability was shown to be good ($\alpha=.85$) (Marteau & Bekker, 1992). Within this study, the internal reliability ranged between .82 and .92 across the 11 timepoints.

To measure curiosity, the Curiosity and Exploration Inventory-II was used (Kashdan et al., 2009). The instrument consists of ten statements, five of which deal with a learner’s motivation to seek out new experiences, and five which assess a learner’s willingness to embrace novel and unpredictable daily life events. Participants were asked to respond to statements on a scale from 1 (*slightly/not at all*) to 5 (*extremely*). An example statement is as follows; “I actively seek as much information as I can in new situations”. No items were reversed scored. The measure has been shown to have good internal reliability ($\alpha=.85$) (Kashdan et al., 2009) within an undergraduate sample. In the current study, internal reliability ranged from .89 to .93 across the 11 timepoints.

Resilience was measured using the Brief Resilience Scale (Smith et al., 2008), which closely aligns statements to the original meaning of resilience (i.e., a person’s ability to rebound from a stressful event). Participants were asked to rate their perceived resilience on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is as follows; “I tend to bounce back quickly after hard times”. Three items were reversed scored. The scale has been tested with a student

population, where Cronbach Alpha scores ranged from .80 to .91 (Smith et al., 2008). For this study, internal reliability ranged from .85 to .92 across the 11 timepoints.

5.2.2.3 *Well-being questionnaire*

The GP-CORE (Evans, Connell, Audin, Sinclair, & Barkham, 2005) was used to evaluate psychological well-being and serves as a shortened form of the CORE-OM. The GP-Core consists of 14 items, with participants asked to rate their perceived behaviours and emotions on a scale from 1 (*not at all*) to 5 (*most/all of the time*). An example item is as follows; “I have felt OK about myself”. Six items were reversed scored. The GP-Core removed all ‘risk’ items which were found in the original Core-OM, and in doing so, was considered more accommodating towards a non-clinical population. When tested with an undergraduate population, the internal reliability was $\alpha = .87$, showing the measure to have strong reliability (Evans et al., 2005). Within this study, the measure had a high internal consistency ranging from .84 to .90 across the 11 timepoint measures. Since its introduction, the measure has been used in several studies investigating well-being in a student population (e.g., Bewick et al., 2010; Cooke et al., 2006).

5.2.2.4 *Experience Sampling Questionnaire*

5.2.2.4.1 *State personality*

State personality was captured using 25 single-item adjectives, each of which mapped onto one of five higher-order traits (i.e., openness, conscientiousness, extraversion, agreeableness, and neuroticism). Participants were asked how each adjective reflected their current thoughts, feelings, or behaviours on a Likert Scale from 1 (*not at all*) to 7 (*completely*). Each dimension consisted of five items, examples of which included: ‘Insecure’ (neuroticism), ‘Curious’ (openness) and ‘Hard-working’ (conscientiousness). Ten of the items were reversed scored, and item order was randomised. The items chosen mirrored adjectives chosen by Fleeson (2001) but also included an extra adjective for each personality dimension.

5.2.2.4.2 *Situational Cues*

Using a framework (PEARLS, see section 2.4) provided by Nofle and Gust (2015), situational cues were measured via a 10-item questionnaire divided across the topics of person, place and task. An example item was ‘How many people are you currently with?’

5.2.2.4.3 *Situational Characteristics*

The ‘Ultra-brief Situational Eight’ instrument (Rauthmann & Sherman, 2016) was used to capture situational characteristics. The instrument is a shortened form of the ‘Situational Eight DIAMONDS’ taxonomy (Rauthmann et al., 2014, see section 2.4). The questionnaire consisted of one item per situation domain and are as follows: Duty (a job or work needs to be done), Intellect (situation requires deep-thinking or demonstrating intellectual capacity), Adversity (being blamed/criticised/threatened for something), Mating (potential romantic partners are present), pOSitivity (situation is potentially enjoyable or pleasant), Negativity (situation is potentially anxiety-inducing or contains negative feelings), Deception (it is possible to deceive someone), and Sociality (social interaction is possible or required). Participants were required to rate their current situation on each of the above domains on a scale from 1 (*not at all*) to 7 (*totally*). The ultra-brief Situational Eight has been used in an undergraduate population, proving to be a valid and reliable instrument. The average convergent reliability was .74 (.86 - .45), indicating that, overall, each of the scales related to the latent construct they were designed to tap into, while discriminant validity was also strong (.37 on average) (Rauthmann & Sherman, 2016).

5.2.2.5 *Linguistic proficiency*

Linguistic proficiency was measured using a C-test, first developed in 1981 by Raatz and Klein Braley (1981). A C-test is a form of reduced redundancy testing whereby elements of a text are deleted and their gaps completed by the test taker. For example, ‘pr_’ becomes ‘proud’ and ‘par_’ becomes ‘parents’ in the following example: “His par____ were pr____.”

For each language tested (French, German, Italian, Spanish), the instrument consisted of four separate paragraphs each on a different topic. These topics differed across languages. Each paragraph contained 25 gaps, and as such, an individual’s score could range between 0 and 100 for each instrument taken. A dichotomous scoring system was used whereby any incorrect answer was marked with 0. To be marked as correct, all linguistic aspects (e.g., spelling, gender marking) of the word had to be correct. Context was provided by the first and last sentence being complete. Tests were not timed and were completed in non-exam conditions.

C-tests have continually been used as an indicator of overall proficiency (Eckes & Grotjahn, 2006), and their validity and reliability tested profusely over the past 50 years. Hastings (2002, p. 24) concluded that “the value of C-testing as a measure of global proficiency in the second language has been demonstrated too many times to be open to dispute”, while Eckes and Grotjahn (2006, p. 292) described C-tests as a “highly reliable, unidimensional instrument.”

The C-tests selected for this study are freely available to the public and are sourced from university language department websites (see Appendix L). While it would be entirely possible for participants to search for and locate these C-tests, given the low stakes nature, there would appear little motivation to actually do so. The same test was used across all learners and timepoints. The main advantage of such an approach is that test difficulty remains constant across the study and as such any improvement in score is most likely due to the test-taker ability as opposed to another test being easier/harder. The limitation, however, is a potential practice effect – by retaking the test, one may naturally get better marks. It is hoped, however, that a gap of more than three months was enough time to minimise this practice effect. Perceived language competencies were also measured at this time in a separate questionnaire.

Perceived competencies were measured via twelve ‘can-do’ statements, rated between 0 (*not at all*) and 10 (*exceedingly well*). Each statement was based on those found on the CEFR website (<https://rm.coe.int/168045b15e>), covering all four language skills (i.e., listening, reading, writing and speaking). An example item is: “I can ... express myself clearly in everyday conversation”.

An overview of instrumentation can be found in Tables 28 to 31 below:

Table 28: Overview of Study Outcomes

Variable	Definition	Measure	Q	I/C?
Outcome variables				
<i>Broad personality trait*</i>	A learner's GENERAL thoughts, feelings and behaviours.	The Big-5 Inventory - 44-items with a 5-point scale (see John & Srivastava, 1999). Captures scores in extraversion, agreeableness, conscientiousness, neuroticism and openness	T	I + C
<i>State personality*</i>	A learner's CURRENT thoughts, feelings and behaviours	Series of 25 adjectives adapted from Fleeson (2001); 7-point rating scale. Captures scores in extraversion, agreeableness, conscientiousness, neuroticism and openness	S	I + C
<i>Perceived monthly level of anxiety*</i>	A self-perceived average rating of how anxious participants had felt over the previous month	6 statements, with a 5-point scale. Followed conventions of the short form State-Trait anxiety Inventory (Marteau & Bekker, 1992)	N	I + C
<i>Perceived monthly level of curiosity*</i>	A self-perceived average rating of how curious participants had felt over the previous month	10 statements with a 5-point scale taken from Kashdan et al. (2009)	N	I + C
<i>Perceived monthly level of resilience*</i>	A self-perceived average rating of how resilient participants had felt over the previous month	6 statements with a 5-point scale taken from The Brief Resilience Scale (Smith et al., 2008)	N	I + C
<i>Perceived monthly well-being*</i>	A self-perceived average rating of participant well-being over the previous month	14 statements with a 5-point scale adapted slightly from Evans et al. (2005)	N	I + C
<i>L2 linguistic proficiency</i>	Level of one's general linguistic proficiency in either French, German, Spanish or Italian	100-item C-test split across four separate topics	L	I
<i>Perceived self-efficacy in L2 ability</i>	Level of self-efficacy in one's L2 ability covering the four skills of language learning.	25-items, scaled between 0-10 based on B2+ Alte can-do statements	L	I

Key: T = broad personality trait; S = state personality; N = narrow personality trait & well-being; L = Linguistic proficiency

* These variables could also serve as predictors in regression models

Table 29: Overview of Programme-specific Variables

Programme-specific variables				
Variable	Definition	Measure	Q	I/C?
<i>Previous Experience</i>	The learner has spent 6+ weeks abroad prior to year abroad	Dichotomous single item YES/NO	D	I
<i>Language Student</i>	The learner was a language major	Dichotomous single item YES/NO	D	I
<i>Two countries</i>	The learner divided their year abroad across two countries	Dichotomous single item YES/NO	D	I
<i>Stayed in Europe</i>	The learner remained in Europe	Dichotomous single item YES/NO	D	I
<i>Comp/Vol</i>	Type of exchange programme undertaken	Dichotomous single item COMPULSORY/VOLUNTARY	D	I
<i>Sojourner Role</i>	Activity undertaking by learner while abroad	Multiple-choice item STUDENT/WORKING/TEACHING	D	I
<i>Length of Stay</i>	Length of stay in months abroad	Open-ended single item	D	I

Key: D = demographic questionnaire

Table 30: Overview of Predictors Associated with Environmental Interpretation²

Environmental perception variables (monthly)				
Variable	Definition	Measure	Q	I/C?
<i>Close bond</i>	The individual has found a new friendship	Dichotomous single item YES/NO	N	I + C
<i>Negative event</i>	The individual experienced a negative event(s)	Dichotomous single item YES/NO	N	I + C
<i>Extracurricular activities</i>	The individual partook in extracurricular activities	Dichotomous single item YES/NO	N	I + C
<i>Travelled</i>	The individual has visited a foreign country	Dichotomous single item YES/NO	N	I + C
<i>Belongingness</i>	The individual has felt welcomed in their environment	Dichotomous single item YES/NO	N	I + C
<i>Feel lonely</i>	The individual has felt predominantly lonely	Dichotomous single item YES/NO	N	I + C

Key: N = narrow personality trait & well-being questionnaire

Table 31: Overview of Variables captured in the Experience Sampling Questionnaire

Situational Cues				
Variable	Definition	Measure	Q	I/C?
<i>People</i>	How many people are the learner with?	Open-ended single item	S	I + C
<i>Place</i>	Where is the learner?	Open-ended single item	S	I + C
<i>Task</i>	What is the learner currently doing?	Open-ended single item	S	I + C
Situational Characteristics				
Variable	Definition	Measure		I/C?
<i>Place security</i>	How secure does an individual feel in their location?	Sliding percentage bar (0-100)	S	I + C
<i>Place familiarity</i>	How familiar is the current location?	Sliding percentage bar (0-100)	S	I + C
<i>Task familiarity</i>	How familiar is the current task?	Sliding percentage bar (0-100)	S	I + C
<i>Task benefit</i>	How beneficial is the current task?	Sliding percentage bar (0-100)	S	I + C
<i>Task enjoyment</i>	How enjoyable is the current task?	Sliding percentage bar (0-100)	S	I + C
<i>Situational Characteristics</i>	A set of 9 situational characteristics describing the current situation	9-items; 7 – point scale adopted from DIAMONDS (Rauthmann et al. (2014)	S	I + C

Key: S = state personality questionnaire/Experience sampling component

² It may also be argued that variables such as *loneliness* and *belongingness* can be associated with the *person*. While I acknowledge this, for the purposes of this study, I have considered them as environmental to make them distinctly different from the captured personality variables. Moreover, it can be argued that the environment does influence the extent a participant feels lonely for example.

5.2.2.6 *Focus groups*

The focus groups were considered as a subsidiary component to the main study and were used as a means of gaining qualitative evidence to explore further the findings of the main study. They have long been used in behavioural sciences and differ from other modes of data collection (e.g., interviews) in that they use group interaction as a direct data collection method (Winke, 2017). In total, three focus groups were conducted, each with an average of five individuals. Group one consisted of purely language learners, while the remaining two groups were a mixture of language and non-language learners. Focus groups were conducted in English, as this was the L1 of both the researcher and participants. For issues considered sensitive, participants were reminded that all responses were anonymous and that they could leave the discussion at any time.

Focus group questions were informed by questions used in focus groups conducted for the LANGSNAP Project (see Tracy-Ventura et al., 2016). Each focus group used the same set of questions which were designed to map on to the key components of the main study. As such, the key constructs which were operationalised in the discussion were the perceived changes in individuals' linguistic and personal development and the perceived environmental factors which were associated with such changes. The questions can be found in Appendix N, and an example question is as follows: "Have you perceived any frustrations with your language development?"

5.2.3 **Study design**

Within Study 2, the primary outcomes were personality development and linguistic development. Each will now be discussed in turn for different study designs were undertaken.

5.2.3.1 *Personality and well-being*

A longitudinal, quasi-experimental, non-equivalent control group design was undertaken to capture data on personality and well-being outcomes. In Study 2, 3rd year student sojourners were considered the treatment group, while those who remained at-home, were considered the (non-equivalent) control group. Consequently, the learning context (abroad/at-home) can be considered the quasi-experimental factor.

An alternative design could have been a within-subjects design, whereby participants receive both the treatment and control condition. While within-subject designs have several advantages over the between-subject design (e.g., requiring fewer participants to achieve the same statistical power and reducing unsystematic variability), they also have several disadvantages, including a

lack of comparison group and numerous threats to internal validity (e.g., history effects, fatigue effects). Within-subject designs have also often been implemented where it is not feasible to implement a mutually exclusive control group, such as in UK SLA orientated studies where study abroad is compulsory (e.g., Rees & Klapper, 2007; Mitchell et al., 2017a). As it was possible to find a mutually exclusive control group to measure change in personality and well-being, it was considered preferable to implement a between-subjects design. Moreover, given that a within-subject design would require measurements to take place both at-home and abroad, data collection would be required for longer than feasible in this project meaning a within-subject design would not be possible.

While a randomised controlled trial (RCT) may be considered the ‘gold-standard’ of evaluative research (Torgerson, C., & Torgerson, D., 2003), in the prism of study abroad research, this design is ethically and morally questionable. It would, after all, not be ethical, to randomly assign students to have the sojourning experiencing, while others must remain in domestic instruction. Moreover, this random assignment may inhibit student well-being of who feel ill-prepared to live abroad, or do not have the financial resources to live abroad. As such, when RCTs are not feasible, the non-equivalent, control group design may be considered the most robust quasi-experimental design from which to warrant causal inference.

Nonetheless, the lack of random assignment does pose issues when inferring causality, for it can result in self-selection bias. This bias exists because a researcher cannot be certain that the two groups are equivalent on all extraneous variables. Therefore, the two groups must be assumed to be non-equivalent (Christensen et al., 2014). As noted by Cohen, Manion and Morrison (2013), the extraneous variables, or confounding variables, can manifest themselves from several sources, including demographic and environmental factors outside the scope of the study. Consequently, the threat to internal validity is increased, for one may reject a hypothesis which was true (Type I error) or accept a false hypothesis (Type II error). One important aspect of the non-equivalent, control group design is the pre-test. Pre-testing is crucial in such a design because it describes how the groups initially compared. As noted by Shadish et al. (2002), the larger the differences at baseline, the stronger the likelihood a selection bias exists. This current study has included a pre-test from which baseline scores have been collected.

One original aspect of the study design in the personality component is the three-tiered hierarchical approach to the personality domains investigated. Broad personality traits follow a

pre, during and post-test design, where trait personality is measured at four timepoints across the year, each with roughly a 10-week gap. This is aligned with much of the study abroad personality research (e.g., Niehoff et al., 2017; Zimmermann & Neyer, 2013) and consequently provides comparability with these studies. Narrow traits, together with well-being, were administered monthly, with the constructs measured in these linking both directly and indirectly to the broader traits. These monthly measures could, therefore, be considered as complementing the broad personality trait measure, just measured more frequently. Lastly, the use of an intensive repeated measurement approach allowed for the detailed capturing of current thoughts, feelings and behaviour. For three, one-week periods, participants were signalled to respond to a brief questionnaire four times per day for one week. This totalled 84 measurement occasions across the study. The use of ESM has been explained in detail in section 2.5, but in brief refers to a technique which enables a researcher to follow a participant over a short period through their daily life, with participants repeatedly notified to complete a brief questionnaire. ESM removes the need for 'pen-and-paper questionnaires' offering the advantage of lower start-up costs, removing geographical boundaries, and enabling easier data collection. Nonetheless, it requires an internet connection and can be inconvenient for participants, given the repeated notifications.

5.2.3.2 *Language*

For the language component of this study, a one-group, pre/post-test design was undertaken. Given the compulsory nature of the year abroad in the UK for language learners, the formation of a control group was not feasible, for many language learners are abroad in their 3rd year of study. Although it was found that York St John learners remain at-home in their 3rd year, the sample available was simply not large enough to be considered representative of language learners across the UK. Attempts were made to recruit participants from York St John, but this resulted in only three participants demonstrating interest. As such, the desire of researching language change in the 3rd year of study in an at-home context was dropped.

Due to the lack of a comparison group, the one-group pre/post-test design must be considered weaker than that of the design implemented to measure personality. This is because it is impossible to ascertain whether these students would have made similar gains if they had remained in formal domestic language instruction. Consequently, the following design does not allow one to evaluate whether sojourning facilitates the development of language learning any more than remaining at-home. Instead, the design can only capture whether sojourners return home (i.e., post year abroad) more proficient than when they began their period abroad.

In sum, the study designs employed offer different research questions to be answered. For personality, the research design is evaluative, capturing differences in changes between the two learning contexts. For the language component, on the other hand, it is purely indicative and describes change across time. It can, as such, provide no inferences concerning the value sojourning has on facilitating language development compared to staying at-home.

5.2.4 Procedure

Figure 3 below details the time frame and sequencing of measures undertaken in this study.

Figure 3: Overview of Data Collection Sequence

Month	Comparison group (remain at-home) n = 70		Treatment group (go abroad) n = 110				
May-June	Piloting		Piloting				
Aug	B	-	B	-	L	-	
Sep	B	N	B	N	-	-	Beginning of 1 st semester – baseline complete
Oct		N		N		-	
Nov	SSSSSS (4x a day for 1 week)	N	SSSSSS (4x a day for 1 week)	N		-	
Dec	B	N	B	N	L	-	
Jan' 19		N		N		-	Beginning of 2 nd semester
Feb	SSSSSS (4x a day for 1 week)	N	SSSSSS (4x a day for 1 week)	N		-	
Mar		N		N		-	
Apr	B	N	B	N		-	Beginning of 3 rd semester
May	SSSSSS (4x a day for 1 week)	N	SSSSSS (4x a day for 1 week)	N		-	
Jun	B	N	B	N	L	-	End of the academic year. YA begins to end
Jul	-		B	-	L	-	
Aug	-		B	-	L	-	
Sep	-		B	-	L	-	Final participant arrives home – YA over
Oct	-					F	Focus group discussion (n = 14)

Key: B: Broad traits
 N: Narrow traits
 S: State personality
 L: Language
 F: Focus Groups

As shown by the timeline, primary data collection commenced in August 2018. After completing the consent form, participants were asked to complete the demographic questionnaire.

Broad personality trait questionnaires were sent every quarter, while narrow trait and well-being questionnaires were sent on the first day of each month. Language instruments were disseminated at the same time as the trait questionnaires at T1, T2, and T4, for those who were language learners. Each learner received a separate C-test for each language they learnt which was then followed by a perceived competencies questionnaire. The experience sampling component of the study which explored state personality and the relevant situational contingencies were administered at three, one-week intervals across the year. In each set week, a participant received four signals a day to respond to a brief questionnaire. The sampling strategy employed was a fixed timing schedule or interval contingent sampling (Wheeler & Reis, 1991). Here, questionnaires were disseminated at 4, 3-hour intervals (9 am, 12 pm, 3 pm, and 6 pm) and participants were asked to describe their current experiences and behaviours at that moment. The advantage of this technique is that participants can “configure their schedules around these reports” (Conner & Lehman, 2012), and they tend to be less burdensome to participants. A limitation of this strategy, however, is that individuals may self-prepare for these reports, and the answers given may not be a true reflection of their stream of consciousness. Given that sojourners were based in various time zones around the world, it was felt that on balance a fixed-time schedule allowed for easier automatised and held greater feasibility for the researcher, in addition to reducing participant burden.

Regarding the focus groups, participation in these took place in October 2019. The focus groups took place in a well-known location and lasted for roughly 45 minutes. Before each focus group, individuals were asked to read a consent form which informed them how the data would be used and how the discussion would proceed. Before commencing, the researcher checked if all individuals were happy to proceed. Each discussion was audio recorded using a SONY handheld recorder which was placed in the middle of the table. After each discussion, the file was transferred via cable to the researcher's encrypted hard drive where the audio was transcribed. No third-party software was used to aid the transcription process. Lastly, transcripts were printed and annotated in order to highlight patterns and key information.

Altogether, the study lasted 16 months, capturing a variety of personality and linguistic measures.

5.2.5 Piloting

As shown in Figure 3 above, piloting took place between May and June, prior to the main study. For the linguistic component, piloting is highly recommended as set out by the validation of C-test guidelines (Grotjahn, 2002). A convenience sampling technique was employed, where

friends and their friends (snowballing) were invited to complete the instruments. In total, eight C-tests per language were tested across three proficiency bands according to the CEFR. For each language tested, there were five individuals in each group. At B1, 1st year language students at Durham and Swansea University were recruited. At B2/C1 level, included individuals had completed a language degree from either Durham or Swansea University. At C2, native speakers of each language were contacted. The purpose of piloting was twofold. Firstly, it provided a space for feedback and editing. Native speakers were particularly helpful in this regard, as some language was considered archaic in the texts and was as such altered. Secondly, by testing across proficiency bands, I was able to rule out texts which were too easy or too hard. Therefore, the texts chosen sufficiently differentiated between proficiency bands. That is to say, there was a significant difference in scores between B1 and B2/C1 levels, and no evidence of ceiling effects in the texts was found. This was important, for their presence, decreases the likelihood “that the testing instrument has accurately measured the intended domain” (Salkind, 2010).

Given that the personality measures were already well-tested within an undergraduate sample, the purpose of piloting these was to receive feedback on software usability and overall level of clarity. For example, one of the state personality adjectives was “perturbable”, which was noted as unfamiliar to the user, and as such dropped from the instrument. The personality measures were tested on the same individuals who undertook the language component at B1 and B2 level.

5.2.6 Ethical considerations

The longitudinal study received ethical approval from the Ethics Board at the Department of Education at Durham University on the 15th of May 2018, while focus group ethical clearance was awarded on the 19th of July 2019. It was ensured that the research followed guidelines provided by both the department and by the British Educational Research Association (2011). The ethical considerations will now be briefly discussed below.

Consent, anonymity and confidentiality

In line with BERA guidelines, all participants who volunteered to partake in the study were well-informed on the study’s purpose and aims before being asked to consent. Once interest had been indicated, the participant received a follow-up email. Attached to this email was an information pack which made clear the purposes of the research, time commitments involved and data usage policy. Once read, the participant was asked to complete the consent form via the

link embedded in the email. It was made explicitly clear that participants could withdraw from the study at any time without consequence or judgement.

Due to the potential sensitivity of the data collected, anonymity was guaranteed. Participants were asked to create an eight alphanumeric code which would serve as their username. In only the demographic questionnaire, were participants asked to provide a contact email so that questionnaires could be sent. Consequently, in no dataset which contained data on personality or language ability, could an email address be linked, as only the Personal ID was displayed to the researcher. Regarding the focus groups, no video recording was undertaken, and anonymity in the study findings upheld for their true names was replaced by pseudonyms.

Lastly, confidentiality was achieved by ensuring that only the researcher could explicitly link an email address with a Personal ID. These email addresses were held securely and encrypted on a hard drive which was only accessible to the researcher. The use of pseudonyms in transcriptions ensured participant confidentiality, and in any instances where names were given in discussions, these have been changed (see Appendix N). All instrument data were stored on the protected server provided by SoSci. Instrument data were only downloaded when needed and was never stored on any computer or hard drive for an extended period.

Data sensitivity and removal

The researcher was acutely aware that for some, both the quantitative and qualitative components of the study may have elicited data which showed extreme negative emotional states, causing discomfort or distress to the research participants. The researcher made clear their responsibilities to participants in such cases, and if required, shared online support networks or took their concerns to an external body. In only one case did this happen, where the above procedure was followed.

As aforementioned, participants could withdraw from the study at any time. In total, 24 participants formally dropped out by email. While a reason was not always given, some spoke of how the requirement to discuss feelings had negative consequences for their mental health. In such instances, their data were securely deleted from all databases.

Experience sampling methodology component

Here, the primary ethical consideration related to the persistent signalling this approach required. This proved bothersome for some individuals, and in two instances, participants emailed the

researcher to signal their desire to be omitted from the measurement mailing list. In both instances, individuals were immediately removed and were sent no more signals. This component also required a time commitment which could interfere with other daily tasks. This was considered, and as aforementioned, the use of a fixed time schedule was implemented to minimise participant burden and aid feasibility on the part of the researcher.

5.2.7 Data analysis

All quantitative data were stored and analysed using R, Excel and SPSS (Versions 24-26). Where necessary, items were reversed scored, and items were averaged across all items which belonged to the same scales. For examples, the ten items corresponding to openness in the Big Five Inventory were averaged to create one openness score at each timepoint.

*5.2.7.1 Multiplicative effect (time * learning context)*

Research questions 2 and 7 looked to determine whether learning context had a stronger or weaker effect on outcome trajectories over time (i.e., a multiplicative effect). As such, a two-level, random intercept, random slope multilevel model approach was undertaken.

Multilevel models can be considered the most appropriate statistical analysis when analysing repeated-measures data and are being increasingly used in behavioural research, replacing the more traditional frequentist procedures (e.g., repeated measures ANOVA). While such analysis has its detractors (e.g., Gorard, 2007; Roli & Monari, 2014), MLMs hold several important advantages over the more traditional techniques. Firstly, MLMs are better equipped at dealing with missing data, and data structures can be unbalanced. In instances of other tests, such as a Repeated Measures ANOVA, casewise deletion is the default option of dealing with missing data, reducing sample size and with-it power and precision of results. Secondly, the assumption of independence can be voided when running an MLM; for example, it may be that specific characteristics are homogenous within-group but differ between-group. This is naturally occurring in longitudinal data for participant responses will likely be similar due to the virtue of being the same individual violating the independence assumption (Bliese, 2006). Traditional statistical analyses such as ANOVAs can produce type 1 errors and biased parameter estimate when this assumption is violated, making multilevel modelling a necessity (Peugh, 2010). Moreover, it is expected that results at T1 and T2 will be more closely related than those at T1 and T4 with this lag one autocorrelation being accepted by MLM. As such, the analyst can use data at both the individual and group level since the standard error can be affected by the

clustered nature of the data (Gorard, 2007). Lastly, MLM enables one to model variability between the covariate and outcome through regression slopes. In other tests such as ANCOVA, it must be assumed that there is homogeneity of regression slopes. Put differently; it is assumed that the relationship between the covariate and the outcome is the same across the different groups (i.e., participants) that make up the predictor variable (Field, 2013). In MLMs, however, it is possible to model this variability explicitly in the regression slopes (Field, 2013).

The equations used for model building were as follows, using trait openness as an example:

Unconditional means model (null model)

As recommended by Nezlek (2001), the first model produced was the null model which contained no predictors. The equation is as follows, with trait Openness being exemplified:

$$\text{Level 1: } \text{openness}_{it} = \beta_{0i} + e_{it}$$

Where t represents the number of different measurement occasions, and i represents the i -th individual ($i = 1 \dots 180$). β_{0i} represents the estimated average openness score, over the specific number of timepoints, for the i -th individual. e_{it} is assumed to be normally distributed serving as the within-person random error. This captures the difference between the observed openness score at time t and the predicted (average) score of the i -th participant (Kwok et al., 2008).

$$\text{Level 2: } \beta_{0i} = \gamma_{00} + u_{0i}$$

Where γ_{00} is the grand mean of the average openness scores for all individual and u_{0i} is the difference between the i -th average openness score and the grand mean. u_{0i} is assumed to be normally distributed with equal variance.

Linear growth model (multiplicative effect)

$$\text{Level 1: } \text{openness}_{it} = \beta_{0i} + \beta_{1i} \text{TIME}_{it} + e_{it}$$

Where β_{0i} is the estimated openness score for the i -th individual at baseline when Time_{it} is equal to 0. β_{1i} is the average monthly change in openness for the i -th individual over time. e_{it} is the within-individual random error and is assumed to be normally distributed with equal variance.

$$\text{Level 2: } \beta_{0i} = \gamma_{00} + \gamma_{01}^{\times(\text{Learning context})}_i + u_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}^{\times(\text{Learning context})}_i + u_{1i}$$

Where γ_{00} is the openness score of the *comparison group* at baseline (Time 0) and γ_{01} how much higher or lower the openness score is at baseline in the *treatment group*. γ_{10} is the change in openness per timepoint for the *comparison group*, and γ_{11} is how much higher or lower the rate of change is in the *treatment group* compared to the *comparison group*. The coefficients u_{0i} and u_{1i} are random effects and capture the random deviations of subjects above or below their group average in terms of intercept and slopes. These random effects are assumed to be normally distributed (Bolger & Laurenceau, 2013).

5.2.7.2 *Capturing the main effect of time*

Research questions 3 and 8 also captured the main effect of time on each outcome, for which random intercept, random slope models were again built. The equations were as follows, using openness as an example:

Unconditional growth model

$$\text{Level 1: } \text{openness}_{it} = \beta_{0i} + \beta_{1i}\text{TIME}_{it} + e_{it}$$

Where β_{0i} is the estimated openness score for the i -th individual at baseline when Time_{it} is equal to 0. β_{1i} is the average monthly change in openness for the i -th individual over time. e_{it} is the within-individual random error and is assumed to be normally distributed with equal variance.

$$\text{Level 2: } \beta_{0i} = \gamma_{00} + u_{0i}$$

$$\beta_{1i} = \gamma_{10} + u_{1i}$$

Where γ_{00} is the average score of openness at baseline and γ_{10} is the average monthly change in openness over the 180 participants. Both u_{0i} and u_{1i} are the between-individual random effects and are assumed to be normally distributed.

5.2.7.3 *Predictors of outcome change and performance*

Research question 4, 10 and 12 each used regression analysis to measure possible predictors of the chosen outcome. For RQ 4, this was broad personality trait change, for RQ 10 this was average well-being scores, and for RQ 12, this outcome variable was language change.

For each question, a preliminary analysis of covariates was first conducted. Here, the dependent variable was regressed on each independent variable, while controlling for the baseline score of the dependent variable, as suggested by Dalecki and Willits (1991). For example, when assessing predictors of language change, baseline proficiency was always controlled for. If the unique

contribution of this variable was significantly different from 0, then this variable was carried forward to the main model.

The final model was built using a stepwise approach. The first model contained the outcome variable. The covariates found significant in the preliminary analysis were then added to the model. Model 2 contained covariates associated with the *person*, Model 3 contained covariates associated with the *environment*, and lastly, Model 4 contained covariates associated with the *exchange programme*. By adding variables into the model this way, it was possible to assess the contribution of variables in predicting the outcome variable and any changes in the variance explained by the addition of these variables.

There is no definitive conclusion regarding the number of subjects per variable (SPV) in order to ensure an accurate prediction. Schmidt (1971) believed that the minimum SPV ranged between 15 and 20, while Green (1991) noted that while a minimum SPV of 20 would be preferable, the minimum SPV required is five. Harrell (2001) argued that the minimum SPV should be set at 10 (Austin & Steyerberg, 2015). Field (2013) notes that an SPV of 10 should suffice if a moderate to large effect is expected. In light of this literature, it was decided that no regression model would exceed more than six predictors.

For each hierarchical regression, all relevant statistical assumptions were tested, and in instances where assumptions were not met, explanations were given. As per Field (2013), the assumptions of a hierarchical regression are as follows: 1) there is independence of observations; 2) linear relationship between the dependent and independent variables both individually and collectively; 3) homoscedasticity of residuals (equal error variance); lack of multicollinearity; 4) no significant outliers and 5) residuals are approximately normally distributed.

Research question 9 used multilevel modelling to detect within and between person differences based on the narrow traits. Using resilience as an example, the level 1 equation is as follows:

$$\text{well-being}_{i} = \beta_{0i} + \beta_{1i}(\text{resilience}_{i}) + e_{i}$$

Where β_{1i} refers to the average resilience score for the i -th individual when the i -th individual is at their typical level of resilience. Resilience has been person-mean centered so that 0 on this variable is the typical resilience score for each individual.

The level 2 equation is as follows: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{resilience}_{i}) + u_{0i}$

here γ_{00} is the grand mean of resilience for the sample, γ_{01} is the effect of the average resilience score for the i -th individual on the sample mean, and μ_{0i} is the well-being mean for each individual AND $\beta_{1i} = \gamma_{10} + \mu_{1i}$ where γ_{10} is the average within-person relationship, and μ_{1i} is the association between resilience and well-being for the i -th individual.

5.2.7.4 *State personality*

Research questions 5 and 6 explored state personality, with the purpose of a) quantifying variability in state personality, b) investigating the systematicity of between-and within-person variability, and c) ascertaining whether state agreeableness could predict a series of outcome variables.

Quantifying variability

For each state (i.e., openness, conscientiousness, extraversion, agreeableness, and neuroticism), a *null model* was built. The model provides information regarding the structure of the data and provides an Intraclass Correlation statistic (henceforth ICC). The ICC value provides a breakdown of how much variability can be found at each level (i.e., between vs within) and also serves as a correlation coefficient. The closer this value is to 1, the stronger the relationship between scores at different timepoints for the same individual.

Systematicity

To explore situational contingencies (see section 2.4), further multilevel analysis was undertaken. A separate multilevel model was carried out for each situational variable, with these being regressed on state agreeableness. Each independent variable at level 1 was mean-person-centred, providing a value which was representative of variation around a participant's own mean level. Using Duty as an example, the level 1 equation is as follows:

$$\text{State agreeableness} = \beta_0 + \beta_1 (\text{Duty}) + e$$

Where β_0 is an individual's mean state agreeableness score across all timepoints. β_1 is an individual's slope for predicting state agreeableness by Duty, and e was the variation of an individual's state agreeableness around their own intercept. The level 2 equation is as follows:

$$\beta_0 = \gamma_{00} + \mu_0$$

$$\beta_1 = \gamma_{10} + \mu_1$$

Where γ_{00} was the grand mean of state agreeableness across all occasions and individuals. μ_0 was the difference between an individual's mean level of state agreeableness and the grand mean. γ_{10} was the grand mean of slope for predicting state agreeableness by Duty, while μ_1 was the deviation of an individual's own slopes from the sample's mean slope.

5.2.7.5 *Language change*

Research question 11 pertained to understanding whether language learners returned home more proficient than when they departed. As the language sample was split between Single Honour learners and Joint Honour learners, a two-way mixed model ANOVA was conducted to ascertain whether the two groups of learners developed at different rates over the year abroad. This data approach allows the testing of a main effect for time (baseline; mid-sojourn; post sojourn) and for Honour programmes (Single Honours & Joint Honours). The analysis also allows for an interaction term. This approach is of particular benefit, given the lack of a control group in the overall study design. It allows tentative conclusions regarding how sojourning may influence language change and provides a little more robustness to the findings.

5.2.7.6 *Qualitative analysis*

Qualitative data were collected to gain deeper insights into how returning sojourners perceived their a) broad traits to change (RQ 4); b) their well-being to fluctuate (RQ 9) and c) their linguistic proficiency to develop (RQ 11).

All qualitative data analysis was conducted in Microsoft Word. Discussions were transcribed verbatim, but marked pauses were not transcribed. Data were analysed using thematic analysis, defined by Nowell, Norris, White and Moules (2017, p. 2) as a method for “identifying, analysing, organising, describing, and reporting themes found within a dataset”. I followed the procedure for thematic analysis, as described by Braun and Clarke (2006). I first familiarised myself with the data by re-reading the transcriptions and noting down initial ideas. Secondly, I highlighted and coded interesting features of the data and then thirdly, collated these codes into potential themes. I then reviewed these themes and identified how these themes fit into the broader purpose of the study. Once the themes were adjudged to fit the broader research questions, they were defined and named (e.g., self-confidence, boredom). Lastly, example quotations were selected and embedded in the text with discussion which related back to the research question.

5.3 Results

Section 5.3 explores the results of Study 2. Attention is first given to the preliminary analysis conducted at baseline.

5.3.1 Preliminary analysis of personality trait data

Attrition

Attrition is not uncommon in longitudinal research, which requires high levels of participant interest and motivation over a sustained period (e.g., Bryman, 2012; Twisk & de Went, 2002). In total, 61 individuals (29 sojourners; 32 non-sojourners) dropped out of the study. Monotonic attrition was even across the two contexts, and there was no clear patterning in the timing of attrition, although attrition was higher in the 2nd half of the study.

The preliminary analysis looked to ascertain the existence of attrition bias. Such bias can occur if people who drop out of the study are systematically different from those who remain (Salkind, 2010). Baseline statistics indicated that those who remained in the study were more open, conscientious, agreeable, emotionally stable and resilient than those who left. The descriptive statistics (Table 32) indicated differences at baseline, which warranted further investigation through a binary logistic regression.

Table 32: Means and SDs of Baseline Personality according to Drop-out Condition

Baseline Broad Trait	Non-drop out (<i>n</i> = 180)		Drop-out (<i>n</i> = 61)	
	Mean	SD	Mean	SD
Openness	3.73	.53	3.66	.59
Conscientiousness	3.64	.60	3.25	.65
Extraversion	3.38	.84	3.40	.80
Agreeableness	3.80	.62	3.59	.66
Neuroticism	3.15	.78	3.20	.88
Baseline Narrow Trait (Aug)				
Anxiety	2.79	.77	3.03	.72
Curiosity	3.47	.79	3.53	.55
Resilience	3.74	.74	3.23	.81

For the purposes of the logistic regression, a newly formed ‘drop-out’ variable (coded as: 0 = remained; 1 = dropped-out) was regressed on each of the baseline traits (see Table 33).

Table 33: Logistic Regression Pertaining to Broad Traits as Predictors of Drop-out

Variable	B	SE	Wald	df	Sig.	Exp (B)	95% C.I. for EXP(B)	
							Lower	Upper
O T1	-.044	.029	2.246	1	.134	.957	.903	1.014
C T1	-.115	.030	14.292	1	<.001	.891	1.050	1.190
E T1	.034	.026	1.697	1	.193	1.034	.983	1.088
A T1	-.040	.029	1.864	1	.172	.961	.908	1.017
N T1	.005	.026	.030	1	.863	1.005	.954	1.057
Constant	4.391	1.935	5.147	1	.023	80.712		

Key: O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism

The regression model was statistically significant as shown by the chi-square test $X^2(5) = 21.925$, $p = .001$, with the model explaining 13.1% of the variance in drop-out and correctly classifying 75.9% cases of drop-out. Sensitivity was 10.2%, specificity was 95%, positive predictive value was 60%, and negative predictive case was 98%. Of the five traits, only baseline conscientiousness showed significance. For clarity, the odds ratio was inverted ($1/.891 = 1.12$), so that for each unit reduction in conscientiousness, the odds of dropping out increased by a factor of 1.12. Conscientiousness has often been found to be predictive of drop-out in personality studies (e.g., Lüdtke et al., 2011; Specht, Egloff and Schmukle, 2011). Narrow trait and well-being baseline scores were all found to be non-significant predictors of drop-out and are therefore not reported in Table 33.

Baseline trait differences (broad and narrow)

Attention is now given to evaluating whether baseline trait differences existed in the 180 individuals who remained in the study for its entire duration. If differences are existent, there is evidence to suggest the presence of selection effects onto a sojourn programme.

It should be noted, that prior to the completion of baseline personality measures, 19 (17%) of the 110 sojourners were already abroad³. This was because these students were volunteering abroad prior to commencement of the formal aspect of the year abroad. Independent t-test analysis indicated that these 19 individuals were significantly more extraverted ($m = 3.83$; $sd = .82$) than the 89 individuals of who were not yet abroad ($m = 3.33$; $sd = .87$), $t(106) = 2.278$, $p = .025$, $d = .57$, which according to Cohen (1988) is a medium-sized effect. These 19 individuals were not significantly different at baseline on any other personality variable measured, and as such, it was decided to retain them for the main analysis.

³ For the remaining 81 students, baseline reflected scores prior to departure on the year abroad

Firstly, analysis was undertaken to identify personality differences at baseline between the at-home sample ($n = 70$) and the sojourning ($n = 110$) sample. The results showed that the two groups differed only in broad trait neuroticism, where sojourners ($m = 3.03$; $sd = .80$) scored on average significantly lower than those at-home ($m = 3.29$; $sd = .77$), $t(176) = 2.185$, $p = .030$. The effect size of this difference was small ($d = .33$), according to Cohen (1988). This finding again justifies only opening the study to 3rd year undergraduates, for had 2nd year undergraduates been invited to participate, it is likely more substantial baseline differences would have been found weakening any potential causal inferences regarding the extent to which learning context influences personality change.

At the heart of understanding self-selection effects, however, is the need to ascertain how different those who volunteered to go abroad were to those who a) remained at-home and b) those for whom the experience was compulsory.

Table 34 displays baseline differences between those who volunteered to sojourn and those who remained at-home. As shown, those who volunteered were found to be significantly less neurotic ($p = .002$, $d = 0.63$) and anxious ($p = .022$, $d = 0.50$), together with being significantly more curious ($p = .026$, $d = 0.50$) and resilient ($p = .002$, $d = 0.67$).

Table 34: Means and SDs of Baseline Personality Traits (Voluntary Sojourners vs At-home Learners)

	Setting	N	Mean/SD	T-test significance level	Effect size (Cohen's d)
Openness	Home	70	3.70 (.52)	$t(106) = .149, p = .882$	0.03
	Abroad	38	3.68 (.56)		
Conscientiousness	Home	70	3.58 (.61)	$t(106) = -1.699, p = .092$	-0.34
	Abroad	38	3.79 (.62)		
Extraversion	Home	70	3.30 (.80)	$t(106) = -.960, p = .339$	-0.19
	Abroad	38	3.46 (.86)		
Agreeableness	Home	70	3.71 (.70)	$t(106) = -.807, p = .422$	-0.16
	Abroad	38	3.83 (.65)		
Neuroticism	Home	70	3.29 (.77)	$t(106) = 3.147, p = .002$	0.63
	Abroad	38	2.79 (.79)		
Anxiety	Home	59	2.86 (.83)	$t(91) = 2.337, p = .022$	0.50
	Abroad	34	2.47 (.68)		
Curiosity	Home	59	3.52 (.76)	$t(91) = -2.268, p = .026$	0.50
	Abroad	34	3.87 (.64)		
Resilience	Home	59	3.44 (.74)	$t(91) = -3.112, p = .002$	0.67
	Abroad	34	3.87 (.60)		

Likewise, when comparing voluntary sojourners against those for whom the experience was compulsory, significant differences were again found (see Table 35). Those who volunteered were shown to be significantly less neurotic ($p = .028$, $d = 0.44$) and anxious ($p = .004$, $d = 0.55$), while being significantly more curious ($p = <.001$, $d = 0.97$) and resilient ($p = <.001$, $d = 0.90$).

Table 35: Means and SDs of Baseline Personality Traits (Voluntary vs Compulsory Sojourners)

	Comp/Vol	N	Mean/SD	T-test significance level	Effect size (Cohen's d)
Openness	Voluntary	38	3.68 (.56)	t (106) = -9.67, p = .336	0.19
	Compulsory	70	3.79 (.52)		
Conscientiousness	Voluntary	38	3.79 (.62)	t (106) = 1.299, p = .197	0.26
	Compulsory	70	3.63 (.59)		
Extraversion	Voluntary	38	3.46 (.86)	t (106) = .357, p = .722	0.07
	Compulsory	70	3.40 (.89)		
Agreeableness	Voluntary	38	3.83 (.65)	t (106) = -.508, p = .613	0.10
	Compulsory	70	3.88 (.52)		
Neuroticism	Voluntary	38	2.79 (.79)	t (106) = -2.231, p = .028	0.44
	Compulsory	70	3.15 (.78)		
Anxiety	Voluntary	34	2.47 (.68)	t (95) = -2.962, p = .004	0.55
	Compulsory	63	2.91 (.71)		
Curiosity	Voluntary	34	3.87 (.64)	t (95) = 4.372, p = <.001	0.97
	Compulsory	63	3.18 (.79)		
Resilience	Voluntary	34	3.87 (.60)	t (95) = 4.230, p = <.001	0.90
	Compulsory	63	3.27 (.70)		

Baseline differences in psychological well-being

Those who volunteered to study abroad, scored significantly higher ($m = 3.93$; $sd = .38$) than both those who were remaining at-home ($m = 3.53$; $sd = .66$; $t (91) = 3.558$, $p = <.001$, $d = 0.76$) and those for whom the experience was compulsory ($m = 3.54$; $sd = .54$; $t (95) = 3.714$, $p = <.001$, $d = 0.79$). According to Cohen (1988) effects were medium to large in size.

5.3.1.1 Summary of findings

The purpose of the preliminary analysis has been to a) to ascertain whether baseline personality traits are predictive of subsequent study drop-out and b) to establish the presence of self-selection effects and understand whether the individuals in each of the learning contexts are significantly different in the outcome variables measured.

These findings have shown that only baseline conscientiousness was predictive of drop-out. Here, it was found that those who were less conscientious at baseline were more likely to drop-

out. Those who volunteered to study abroad also demonstrated several significant differences at baseline indicating self-selection. Those who volunteered to study abroad were less neurotic, less anxious, more curious and more resilient than those who chose to remain. These differences also remained when comparing individuals for whom the study abroad was compulsory with those for whom the experience was voluntary.

5.3.2 Understanding the role of learning context in broad and narrow trait change

Section 5.3.2 explores whether learning context has a direct effect on accounting for differences in *broad* (e.g., openness, conscientiousness, and extraversion) and *narrow* traits (i.e., anxiety, curiosity and resilience) trajectories over time. The analysis was guided by the following question and examined data collected from both sojourners and non-sojourners:

RQ2: Do sojourners experience significantly greater personal growth as measured by broad and narrow personality traits compared to non-sojourners?

Table 36 presents the descriptive statistics for each of the five traits at each timepoint. Scores over time indicated *sojourners* became more open, extraverted and agreeable, while also becoming less neurotic. Conscientiousness displayed little change. *Non-sojourners* were found to become more conscientious and extraverted, while also becoming less agreeable. Openness and neuroticism showed little change.

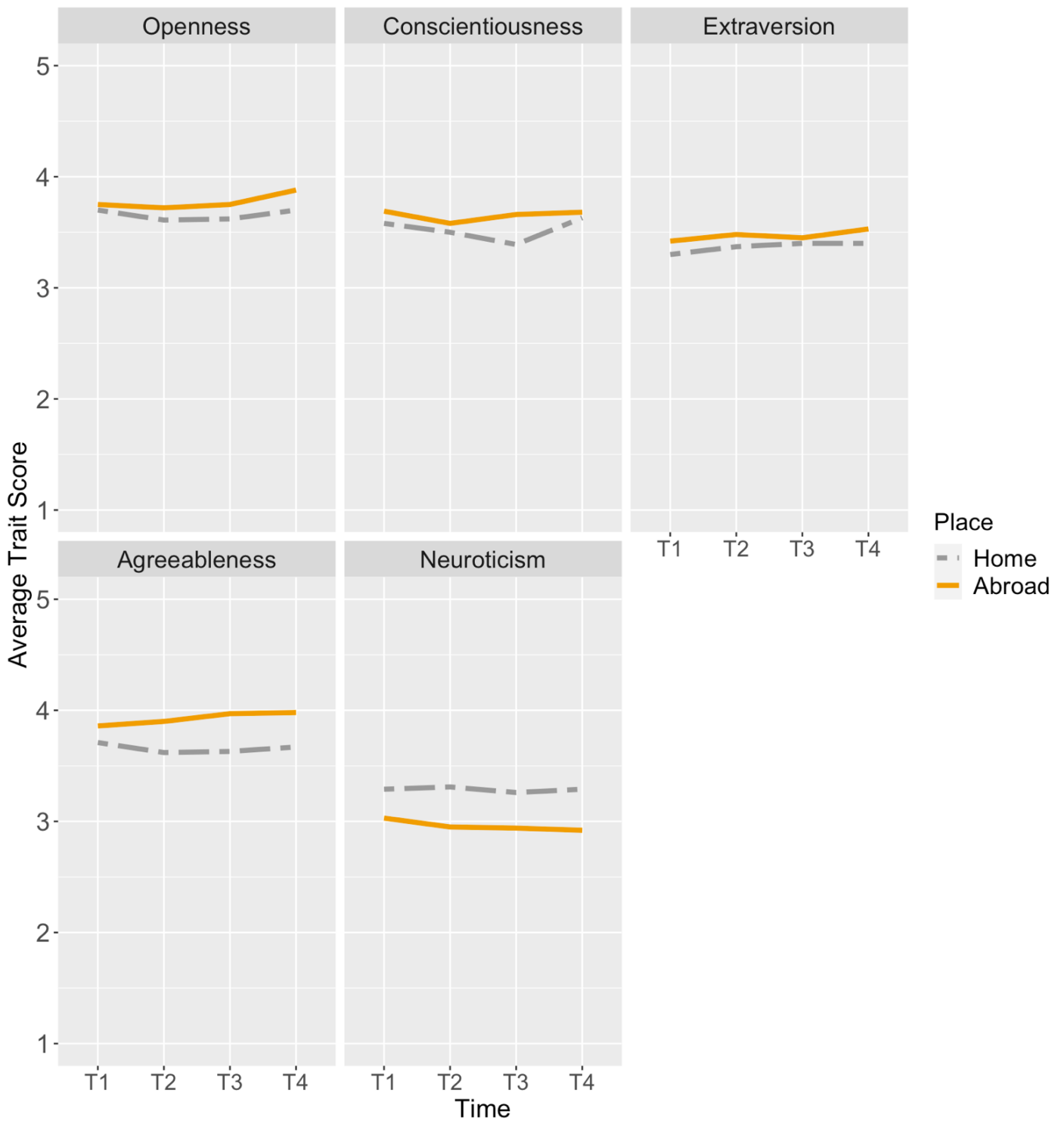
Table 36: Means and SDs of Broad Personality Traits across Time according to Learning Context

		Abroad			Home		
		N	Mean	SD	N	Mean	SD
O	T1	108	3.75	.53	70	3.70	.52
	T2	109	3.72	.60	68	3.61	.55
	T3	107	3.75	.63	68	3.62	.55
	T4	102	3.89	.59	65	3.70	.52
	T4 - T1	100	.103 (Min: -.80; Max: 1.20)	.39	65	-.003 (Min: -1.10; Max: 1.00)	.37
C	T1	108	3.69	.60	70	3.58	.61
	T2	109	3.58	.64	68	3.50	.65
	T3	107	3.66	.65	68	3.50	.64
	T4	102	3.69	.68	65	3.63	.63
	T4 - T1	100	-.018 (Min: -1.66; Max: .89)	.39	65	.063 (Min: -.60; Max: 1.00)	.33
E	T1	108	3.42	.88	70	3.30	.80
	T2	109	3.48	.84	68	3.37	.79
	T3	107	3.45	.80	68	3.39	.79
	T4	102	3.54	.82	65	3.40	.80
	T4- T1	100	.122 (Min: -1.62; Max: 2.25)	.60	65	.120 (Min: -0.90; Max: 1.60)	.47
A	T1	108	3.86	.56	70	3.71	.70
	T2	109	3.90	.58	68	3.62	.75
	T3	107	3.96	.55	68	3.63	.73
	T4	102	4.00	.54	65	3.67	.78
	T4- T1	100	.111 (Min: -1.44; Max: 1.33)	.47	65	-.023 (Min: -1.30; Max: .80)	.43
N	T1	108	3.03	.80	70	3.29	.77
	T2	109	2.95	.80	68	3.31	.88
	T3	107	2.94	.85	68	3.26	.76
	T4	102	2.90	.88	65	3.29	.72
	T4- T1	100	-.155 (Min: -2.75; Max: 2.00)	.77	65	-.007 (Min: -1.10; Max: 1.60)	.54

Note: O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism

To ascertain the value of undertaking a sojourn programme, the first analysis looked to establish the existence of an interaction effect between time and learning context on broad trait change. Before conducting the main analysis, the data were graphically inspected and presented in Figure 4. The plots suggest slight deviations from linearity and that across time, sojourners tended to score progressively higher than non-sojourners in trait openness and trait agreeableness.

Figure 4: Broad Trait Interaction Plots Between Time and Learning Context



The simplest model to capture change over time is a two-level, random intercept, random slope model which estimates the average trait trajectory in both learning contexts. The model treats the learning context dummy (0 = at-home; 1 = abroad) as both a main effect and as an interaction with time. Within this model, the average change trajectory in the ‘at-home’ group is estimated directly, while in the ‘abroad’ group, this trajectory is estimated indirectly (Bolger & Laurenceau, 2013; Leckie, 2019). One assumption of such a model is that change is linear over time. It should be noted, that as shown in Figure 4, there are small deviations from linearity, but nonetheless, these models are a good place to begin analysing the data.

Table 37: Parameter Estimates for Linear Growth Model of Broad Traits as a Function of Learning Context

	Trait O		Trait C		Trait E		Trait A		Trait N	
	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2
Fixed effects										
Intercept	3.72*	3.67*	3.62*	3.53*	3.44*	3.31*	3.83*	3.68*	3.07*	3.29*
	(0.03)	(0.06)	(0.04)	(0.07)	(0.05)	(0.09)	(0.04)	(0.07)	(0.05)	(0.09)
Time		-.002		0.01		0.04		-.008		-0.01
		(0.01)		(0.01)		(0.02)		(0.01)		(0.02)
Learning Context		0.05		0.11		0.11		0.18		-0.27*
		(0.08)		(0.09)		(0.12)		(0.09)		(0.11)
T * LC		0.03		-0.01		<-.001		0.05*		-0.03
		(0.02)		(0.02)		(0.02)		(0.02)		(0.03)
Random effects										
Intercept	0.25*	0.22*	0.33*	0.28*	0.56*	0.61*	0.32*	0.29*	0.52*	0.49*
	(0.50)	(0.47)	(0.57)	(0.53)	(0.75)	(0.78)	(0.55)	(0.54)	(0.72)	(0.70)
Time		-		<.001*		0.01*		<.001*		0.01*
				(0.03)		(0.11)		(0.03)		(0.13)
Residual	0.07*	0.09*	0.08*	0.08*	0.11*	0.08*	0.09*	0.10*	0.17*	0.14*
	(0.27)	(0.30)	(0.29)	(0.28)	(0.33)	(0.29)	(0.33)	(0.33)	(0.41)	(0.37)
ICC	0.77		0.79		0.83		0.77		0.75	
Model fit statistics										
-2 Log Likelihood	645.61	616.02	755.53	744.74	993.69	958.96	1206.66	1169.67	645.61	616.02
<i>Key: O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism</i>										
<i>Note: Level 1 model: $\text{trait}_{it} = \beta_{0i} + \beta_{1i}(\text{TIME})_{it} + e_{it}$, where β_{1i} refers to the average change in a broad trait score for the i-th individual over time. Level 2 model: $\beta_{0i} = \gamma_{00} + \gamma_{01}^X(\text{LEARNING CONTEXT})_i + \mu_{0i}$ where γ_{00} is the average score of the trait at the initial timepoint and γ_{01} is how much higher or lower the trait score is at baseline in the <i>treatment group</i> AND $\beta_{1i} = \gamma_{10} + \gamma_{11}^X(\text{LEARNING CONTEXT})_i + \mu_{1i}$ where γ_{10} is the change in the trait per timepoint for the <i>comparison group</i> and γ_{11} is how much higher or lower the rate of change is in the <i>treatment group</i> compared to the <i>comparison group</i>.</i>										
* $p < .05$										

Table 37 presents two models for each trait. In *Model 1 (null model)*, the intercept serves as the average trait score across all individuals and timepoints. Treating agreeableness as an example, each participant was estimated to score $\beta_0 = 3.83$ on average across all timepoints. The model

also reports the ICC value (.77). As such, 77% of the variation in agreeableness scores was found at the between-person level, and 23% was at the within-person level. This served as evidence that individual differences existed within the sample. It also indicated that at any two timepoints, agreeableness scores for the same individual were highly correlated ($r = .77$), which given the high stability found at the broad trait level, is to be expected. Consequently, there was evidence of rank-order stability in the traits. Put differently, those who scored higher than others on a particular trait at one timepoint tended to score higher than others at other timepoints.

Model 2 served as the *linear growth model*. Treating agreeableness as an example again, the model specified the average baseline score for non-sojourners to be $\beta_0 = 3.68$ and that for each point in time, the average non-sojourner became less agreeable by $\beta_1 = -.008$ units. Sojourners were estimated at baseline to be on average more agreeable than non-sojourners ($\beta_2 = 0.18$). Most importantly, to the research question itself, the interaction term between time and learning context was found to be significant ($\beta_3 = 0.05, p = .03$). This coefficient estimated that at each point in time, the average sojourner scored 0.05 units higher in agreeableness than non-sojourners. This provided evidence that sojourning can facilitate the change in trait agreeableness more than if remaining at-home. The model predicted that the average sojourners scored 0.23 units higher in agreeableness at the end of the academic year than the average non-sojourner. Regarding the random coefficients, the intercept variance has remained unchanged, indicating that learning context explained little variation (2.15%) in agreeableness scores at baseline. Conversely, the slope variation reduced by 10%, indicating that 10% of the variation in trait scores over time was attributable to the learning context. The residual variance was virtually unchanged, which was not surprising given that learning context is an individual-level covariate and as such cannot explain within-person variance (Leckie, 2019).

The final analysis ascertained whether the significant interaction term for trait agreeableness remained after controlling for the measured personality variables. First, the broad traits (i.e., openness, conscientiousness, extraversion and neuroticism) were controlled for, where a significant interaction remained ($\beta_7 = 0.04, p = .042$). Similarly, baseline narrow traits (i.e., anxiety, curiosity and resilience) were controlled for in a separate model, where a significant interaction term was again found ($\beta_6 = 0.05, p = .039$). Consequently, regardless of whether trait personality (broad and narrow) was the same across learning contexts at baseline, sojourners were still predicted to become more agreeable over time compared to non-sojourners, providing

stronger evidence that it was the sojourning context itself, which facilitated agreeableness growth.

Concerning the remaining broad traits, Table 37 indicated no further significant interaction terms between learning context and time. Growth trajectories were not estimated to be significantly different across learning contexts. The interaction terms for the remaining traits are as follows: openness ($\beta_3 = 0.03, p = .09$), conscientiousness ($\beta_3 = -0.01, p = .46$), extraversion ($\beta_3 = <-.001, p = .76$) and neuroticism ($\beta_3 = -0.03, p = .37$).

5.3.2.1 *Capturing differing narrow trait trajectories between learning contexts*

Turning attention to the narrow traits. Table 38 presents the descriptive statistics for each of the three narrow traits across the year. Scores over time indicated *sojourners* became less anxious and more curious over time, while *non-sojourners* fluctuated in their anxiety over time while becoming less curious. Learners in each learning context demonstrated fluctuations in resilience. This patterning is captured in Figure 5, which displays the interaction plots for each narrow trait.

Table 38: Means and SDs of Narrow Personality Traits across Time according to Learning Context

		Abroad			Home		
		N	Mean	SD	N	Mean	SD
Anxiety	Aug	97	2.75	.73	59	2.86	.83
	Sept	104	2.66	.83	68	2.93	.83
	Oct	103	2.65	.72	67	3.06	.73
	Nov	107	2.61	.73	69	3.16	.80
	Dec	109	2.65	.77	69	2.98	.80
	Jan	104	2.52	.76	67	2.96	.85
	Feb	108	2.57	.83	68	2.99	.69
	Mar	108	2.51	.78	63	3.01	.82
	Apr	105	2.42	.74	56	2.99	.90
	May	105	2.41	.74	61	3.18	.85
	June	100	2.31	.68	61	2.49	.85
Curiosity	Aug	97	3.42	.81	59	3.52	.76
	Sept	104	3.54	.72	68	3.37	.83
	Oct	103	3.60	.77	67	3.21	.87
	Nov	107	3.55	.85	69	3.17	.93
	Dec	109	3.37	.81	69	3.14	.93
	Jan	104	3.48	.81	67	3.07	1.04
	Feb	108	3.55	.79	68	3.10	.88
	Mar	108	3.61	.86	63	3.04	.98
	Apr	105	3.65	.74	56	3.02	.86
	May	105	3.65	.85	61	2.98	1.02
	June	100	3.72	.82	61	3.48	.91
Resilience	Aug	97	3.45	.76	59	3.41	.74
	Sept	104	3.69	.76	68	3.32	.96
	Oct	103	3.59	.78	67	3.37	.76
	Nov	107	3.71	.78	69	3.11	.85
	Dec	109	3.57	.82	69	3.32	.89
	Jan	104	3.55	.74	67	3.22	.86
	Feb	108	3.56	.82	68	3.31	.74
	Mar	108	3.54	.77	63	3.34	.81
	Apr	105	3.65	.79	56	3.27	.82
	May	105	3.60	.79	61	3.22	.92
	June	100	3.70	.78	61	3.49	.80

Figure 5: Narrow Trait Interaction Plots Between Time and Learning Context

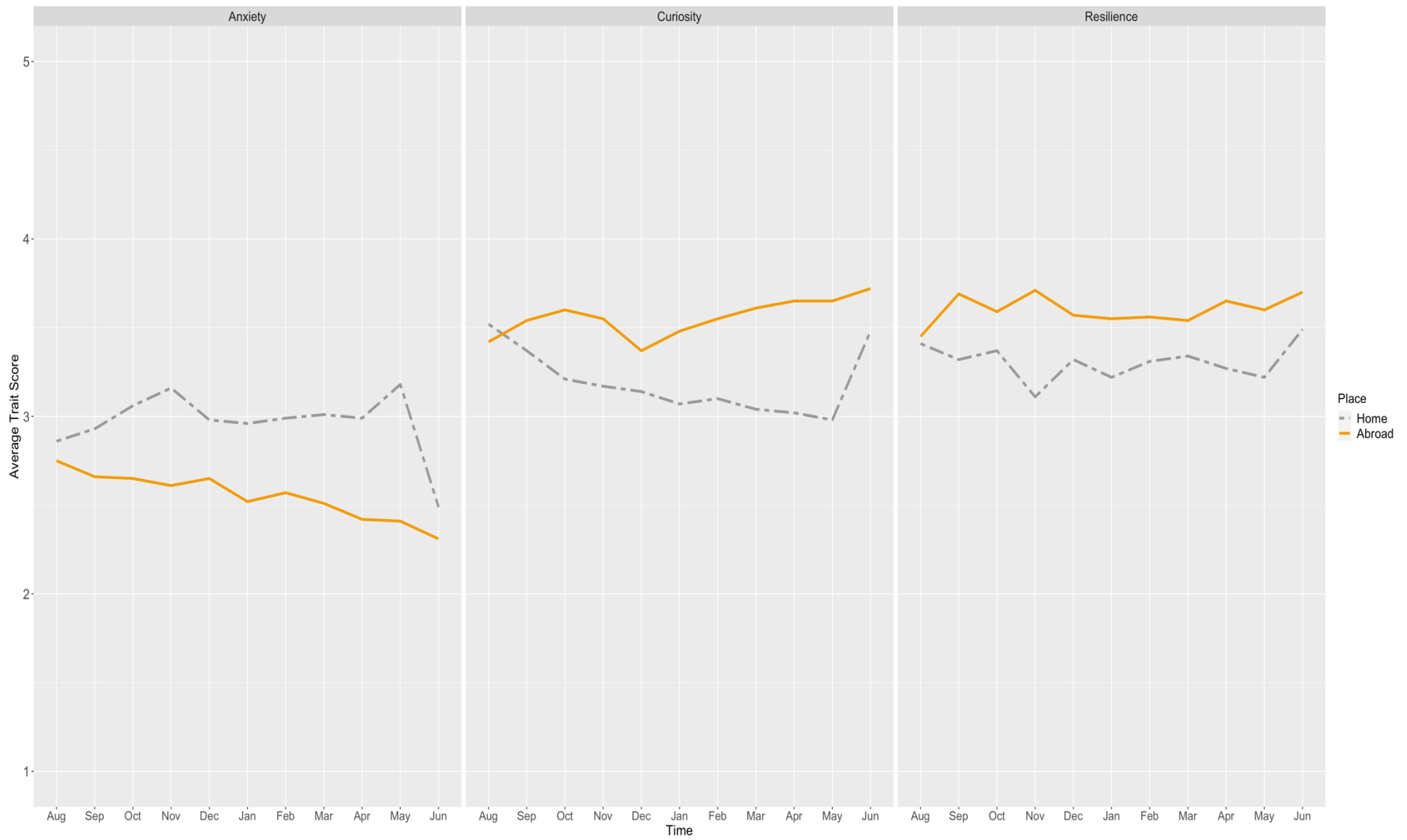


Table 39: Parameter Estimates for Linear Growth Model of Broad Traits as a Function of Learning Context

	Trait Anxiety		Trait Curiosity		Trait Resilience	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Fixed effects						
Intercept	2.72* (0.04)	3.01* (0.08)	3.42* (0.05)	3.30* (0.08)	3.49* (0.04)	3.33* (0.08)
Time		-0.01 (.009)		-0.01 (.00)		-.001 (.008)
Learning Context		-0.27* (0.10)		0.16 (0.11)		0.23* (0.10)
T * LC		-0.02* (0.01)		0.03* (0.01)		.009 (0.01)
Random effects						
Intercept	0.34* (0.58)	0.32* (0.57)	0.42* (0.65)		0.32* (0.57)	0.30* (0.55)
Time		.001* (0.04)	0.34* (0.58)			.001* (0.03)
Residual	0.31* (0.56)	0.30* (0.55)	.55		0.33* (0.57)	0.33* (0.57)
ICC	.52				.49	
Model fit statistics						
-2 Log Likelihood	3576.41	3389.27	3763.42	3631.53	3651.16	3539.21

Note: Level 1 model: $\text{trait}_{it} = \beta_{0i} + \beta_{1i}(\text{TIME})_{it} + e_{it}$, where β_{1i} refers to the average change in a broad trait score for the i -th individual over time. Level 2 model: $\beta_{0i} = \gamma_{00} + \gamma_{01} \times (\text{LEARNING CONTEXT})_i + \mu_{0i}$ where γ_{00} is the average score of the trait at the initial timepoint and γ_{01} is how much higher or lower the trait score is at baseline in the *treatment group* AND $\beta_{1i} = \gamma_{10} + \gamma_{11} \times (\text{LEARNING CONTEXT})_i + \mu_{1i}$ where γ_{10} is the change in the trait per timepoint for the *comparison group* and γ_{11} is the difference in rate of change between the two groups.

* $p = <.05$

Anxiety

Table 39 presents the output of the multilevel model conducted for narrow trait anxiety. Model 1 served as the *null model*, indicating the average anxiety score across all timepoints and individuals to be $\beta_0 = 2.72$. Model 2 served as the full model and can be interpreted as follows:

Non-sojourners were estimated to have a baseline anxiety score of $\beta_0 = 3.01$, and at each timepoint, the average *non-sojourner* was estimated to become less anxious by $\beta_1 = 0.01$ units. Sojourners were estimated to score $\beta_2 = 0.27$ units lower at baseline, with the average sojourner having an estimated anxiety score of 2.74. The interaction term between time and leaning context was significant and estimated that at each timepoint, sojourners scored $\beta_4 = .02$ units lower in anxiety than non-sojourners⁴. Regarding the random part of the model, intercept variance, slope variance, and residual variance remained relatively unchanged. Learning context, therefore, explained little variation at the within-person level. Consequently, it can be said that sojourning fostered an accelerated decrease in anxiety over time compared to remaining at-home.

⁴ When taking into account only August to May, the interaction term remained significant ($p = <.001$), with *sojourners* estimated to score $\beta_4 = .05$ units lower in anxiety at each timepoint than *non-sojourners*.

Serving as complementary analysis, further multilevel models were run in order to ascertain whether the interaction term between time and learning context remained after controlling for variables captured in the study. Firstly, the five broad trait baseline scores were controlled for, where a significant interaction term remained ($\beta_9 = -0.02, p = .03$). Secondly, the *environmental* variables (e.g., loneliness) were controlled for (in a separate analysis), where a significant interaction term was found ($\beta_{10} = -0.02, p = .02$). The respective narrow traits (i.e., resilience and curiosity) were last to be controlled for, where the interaction term became non-significant ($\beta_5 = -0.01, p = .17$)⁵. This finding suggested that when resilience and curiosity scores were held constant, the anxiety trajectories of each group were no longer significantly different. This finding may be an outcome of the strong association between anxiety and resilience, whereby more resilient learners tend to be less anxious (Charney, 2003; Haddadi & Besharat, 2010).

Curiosity

Table 39 presents the output of the multilevel model conducted for curiosity. All individuals, across all timepoints, were estimated to have an average curiosity score of $\beta_0 = 3.42$ (Model 1). At baseline, *non-sojourners* were estimated to have a baseline curiosity score of $\beta_0 = 3.30$, and at each timepoint, the average *non-sojourner* was estimated to become less curious by $\beta_1 = 0.01$ units. At baseline, *sojourners* were estimated to score $\beta_2 = 0.16$ units higher than *non-sojourners* with an estimated curiosity score of 3.46. The interaction term between time and learning context demonstrated significance. Here, the model estimated the average *sojourner* to score $\beta_4 = 0.03$ ($p = <.001$) units higher in curiosity at each timepoint than *non-sojourners*⁶. This interaction term indicates a direct effect of sojourning on perceived curiosity and suggests that those who go abroad become significantly more curious over time compared to those who remain at-home. The residual variance is relatively unchanged, and this is to be expected given that treatment status cannot explain individual variation as it is an individual-level covariate.

Three further multilevel models were run, with the first controlling for baseline broad traits ($\beta_9 = -0.02, p = .03$), the second for environmental variables ($\beta_{10} = -0.02, p = .03$) and the third for monthly narrow trait scores ($\beta_5 = -0.02, p = .03$). In each of the models, the interaction term remained significant, indicating that after controlling for the captured variables, those abroad became on average more curious than those who remained at-home.

⁵ When taking into account only August to May, the interaction term for anxiety remained significant ($\beta_5 = -0.03, p = <.001$) after holding resilience and curiosity scores constant.

⁶ When taking into account only August to May, the interaction term for curiosity remained significant ($p = <.001$), with *sojourners* estimated to score $\beta_4 = .06$ units higher in curiosity at each timepoint.

Resilience

Table 39 presents the output of the multilevel models conducted for resilience. As per earlier models, Model 1 estimates the average resilience score across all timepoints and individuals to be $\beta_0 = 3.42$. Model 2 estimates the average *non-sojourner* to have a baseline score of $\beta_0 = 3.33$ and demonstrate a reduction in resilience by $\beta_1 = 0.01$ units per timepoint. The model estimates *sojourners* to be more resilient at baseline by $\beta_2 = 0.23$. The interaction term ($\beta_4 = .009$) was shown to be non-significant, indicating *sojourners* did not become more resilient than *non-sojourners* over time⁷. The intercept variance, slope variance and residual variance have all remained virtually unchanged, suggesting learning context explained little variation at the within-person level.

The final step of the analysis controlled for measured variables in order to establish whether the interaction between time and learning context remained non-significant when controlling for these. Three models were run, which accounted firstly, for baseline traits ($\beta_9 = 0.01, p = .40$), secondly, for environmental predictors ($\beta_{10} = .008, p = .46$) and lastly, for the narrow traits ($\beta_5 = -0.01, p = .19$). As such, even after controlling for accompanying variables, *sojourners* did not score significantly higher in resilience over time than those who remained at-home.

5.3.2.2 *Summary of findings*

The purpose of section 5.3.2 has been to identify whether sojourning has a direct effect on the trait trajectories of both the broad and narrow traits.

The findings have shown that the average *sojourners* became significantly more agreeable and more curious as a result of sojourning. Due to the inclusion of a comparison group, it can be said that learning context is an underlying causal mechanism in accounting for trait change. There was also evidence to indicate that *sojourners* became significantly less anxious over time compared to those at-home. This effect did, however, disappear when controlling for resilience. As such, evidence regarding the influence of study abroad on anxiety change cannot be considered as conclusive as that of agreeableness or curiosity but does nonetheless indicate that sojourning may have had a direct effect on perceived anxiety.

⁷ When taking into account only August to May, the interaction term remained non-significant ($p = .15$), with *sojourners* estimated to score $\beta_4 = .02$ units higher in resilience at each timepoint.

5.3.3 Assessing the main effect of time in broad trait trajectories

Section 5.3.3 looks to ascertain whether broad trait personality changes significantly in sojourners over time and whether these trajectories can be considered significantly similar. Given the lack of significant interaction effects in section 5.3.2, it can be assumed that the main effect of time for the sojourners sample would not be sufficiently different from that of non-sojourners. Therefore, analysis focuses on the sojourners sample only and is guided by the following research question and examined data for sojourners only:

RQ3: Do sojourners experience significant broad trait change over time, and is this uniform across all learners?

Multilevel models were built to capture individual trajectories in the personality variables over time. Graphical inspection of individual plots indicated that participants not only began with differing trait scores but also varied in their trajectories. As such, regardless of learning context, there were substantial individual differences in directionality and steepness of trait trajectories. This served as evidence of differing rates of growth, and while some displayed stronger trait-like tendencies over the year, others weakened in such like tendencies. These differences were captured using a random intercept, random slope model, where trait score was regressed on a linear time trend. This model allows for individuals to follow to their own linear trajectory with a unique intercept (starting point) and slope (scores over time) (Gelman & Hill, 2007).

With each multilevel model, the *null model* and the *conditional growth model* are presented (Table 40). The Null Model (*Model 1*) presents the average trait score across all timepoints for all sojourners, with the model predicting the average sojourner to have an openness score of $\beta_0 = 3.76$ across all four timepoints for example. The ICC values ranged between 0.69 (agreeableness) and 0.82 (extraversion), indicating that between 69% and 82% of the variation in trait scores was found at the between-person level, while between 18% and 31% of the variation was at the within-person level. Scores at any two points in time were also highly correlated ($r = .69 - .82$).

Model 2 is the *conditional growth model*. Here, the fixed effect of *Time* estimates sojourners to become significantly more open ($\beta_1 = 0.03, p = 0.01$), agreeable ($\beta_1 = 0.04, p = .006$) and less neurotic ($\beta_1 = -0.04, p = 0.04$) over time. For example, the model estimates that the average sojourner departs with an agreeableness score of $\beta_0 = 3.87$ and arrives home with a score of $\beta = 4.03$ ($0.04 * 4 + 3.87$). For agreeableness, the random coefficients indicate an intercept variance (σ_{u0}^2) of 0.23 and a slope variance (σ_{u1}^2) of .007. As noted in Leckie (2019), these scores can be interpreted by calculating a range of intercepts which represent the middle 95% of individuals in the population. Consequently, it is predicted that 95% of sojourners begin the study with an agreeableness score between 2.94 and 4.79 and that 95% sojourners will score between -0.11 and

0.19 higher or lower across time. Therefore, some individuals were estimated to become less agreeable during the year abroad, while others became more agreeable, representing substantial individual differences in change over time. Lastly, the student residual variance (σ^2_e) was estimated to be 0.08, predicting that at any given timepoint, students will score roughly half a unit (0.54) above or below their own trajectory. Table 41 presents the range of intercept, slope and residual variances in which one would expect to find the middle 95% of sojourners in the population for each trait. Like agreeableness, there are considerable differences in individual slope trajectories over time, as demonstrated by the slope variance column. For each trait, some individuals saw a weakening of trait-like responses over time, while others strengthened in these trait-like responses, and further investigation is warranted as to why this may be (section 5.3.4).

Table 40: Parameter Estimates for Linear Growth Model of Broad Traits across Time (Sojourners)

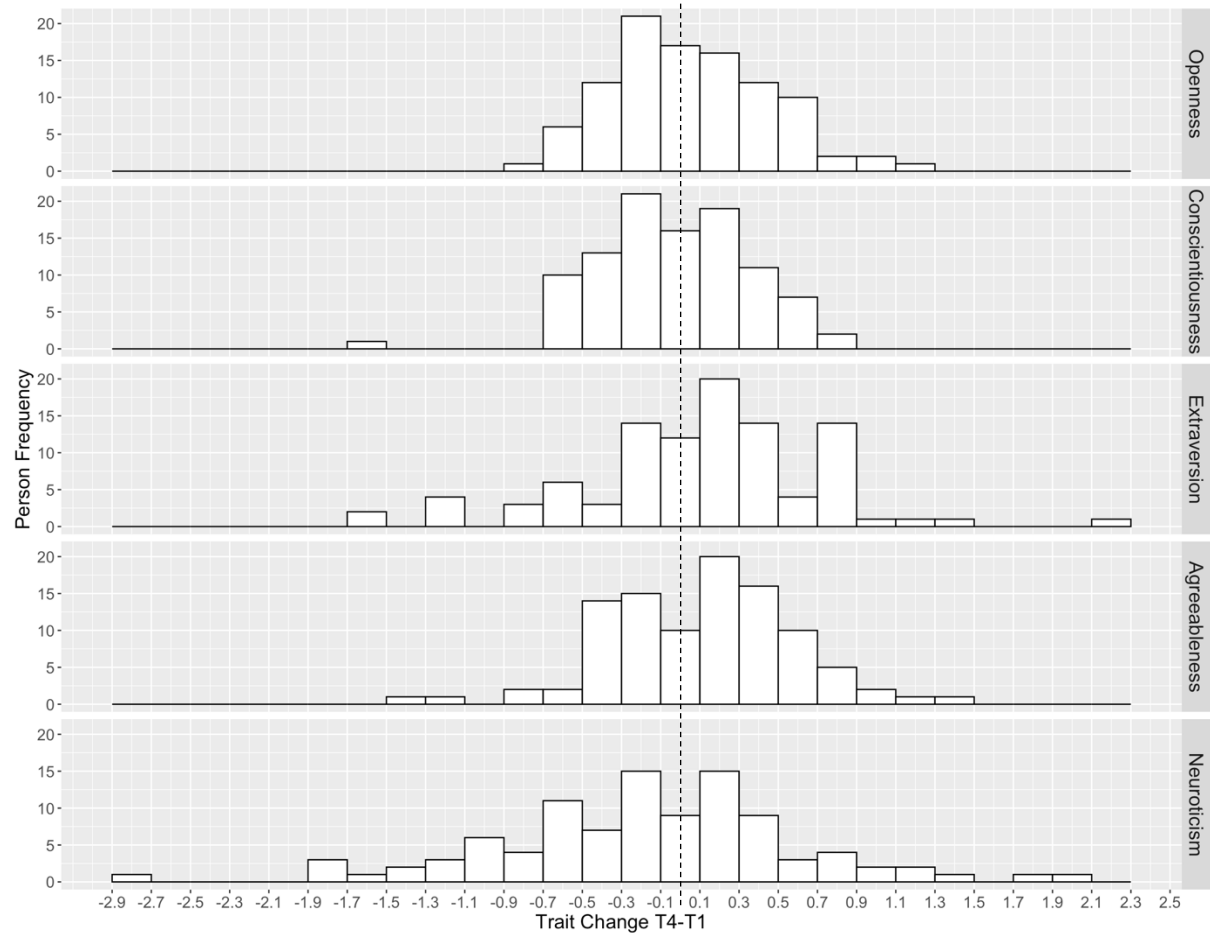
	Trait O		Trait C		Trait E		Trait A		Trait N	
	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2
Fixed effects										
Intercept	3.76*	3.72*	3.65*	3.65*	3.48*	3.43*	3.93*	3.87*	2.95*	3.02*
	(0.05)	(0.05)	(0.07)	(0.05)	(0.07)	(0.07)	(0.04)	(0.05)	(0.07)	(0.07)
Time		0.03*		.001		0.03		0.04*		-0.04*
		(0.01)		(0.01)		(0.01)		(0.01)		(0.02)
Random effects										
Intercept	0.27*	0.23*	0.33*	0.28*	0.57*	0.52*	0.21*	0.23*	0.49*	0.48*
	(0.52)	(0.45)	(0.58)	(0.53)	(0.72)	(0.72)	(0.46)	(0.44)	(0.70)	(0.69)
Time		.001*		.001*		0.01*		.007*		0.03*
		(0.03)		(0.03)		(0.13)		(0.08)		(0.17)
Residual	0.07*	0.09*	0.08*	0.07*	0.12*	0.17*	0.09*	0.08*	0.19*	0.14*
	(0.27)	(0.31)	(0.29)	(0.28)	(0.35)	(0.41)	(0.30)	(0.29)	(0.44)	(0.38)
ICC	0.78		0.79		0.82		0.69		0.71	
Model fit statistics										
-2 Log LH	405.76	375.38	466.95	459.16	646.50	627.05	461.23	439.79	775.05	740.89
Key: O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism										
Note: Level 1 model: $\text{trait}_{it} = \beta_{0i} + \beta_{1i}(\text{TIME})_{it} + e_{it}$, where β_{1i} refers to the average change in a broad trait score for the i-th individual over time. Level 2 model: $\beta_{0i} = \gamma_{00} + u_{0i}$ and $\beta_{1i} = \gamma_{10} + u_{1i}$, where γ_{00} is the average score of the trait at the initial timepoint and γ_{10} is the average monthly change in the trait. * $p < .05$										

Table 41: Descriptive Statistics for Broad Trait Slopes Estimated by MLM (Sojourners)

Trait	Intercept Variance	Slope Variance	Residual Variance
<i>Openness</i>	$2.79 \leq \pi_1 \leq 4.64$	$-0.02 \leq \pi_1 \leq 0.08$	$-0.58 \leq \pi_1 \leq 0.58$
<i>Conscientiousness</i>	$2.63 \leq \pi_1 \leq 4.66$	$-0.05 \leq \pi_1 \leq 0.05$	$-0.50 \leq \pi_1 \leq 0.50$
<i>Extraversion</i>	$2.01 \leq \pi_1 \leq 4.12$	$-0.04 \leq \pi_1 \leq 0.06$	$-0.80 \leq \pi_1 \leq 0.80$
<i>Agreeableness</i>	$2.94 \leq \pi_1 \leq 4.79$	$-0.11 \leq \pi_1 \leq 0.19$	$-0.54 \leq \pi_1 \leq 0.54$
<i>Neuroticism</i>	$1.66 \leq \pi_1 \leq 4.37$	$-0.37 \leq \pi_1 \leq 0.29$	$-0.72 \leq \pi_1 \leq 0.72$

The presence of individual differences can also be demonstrated graphically through histograms (Figure 6). These histograms display the mean change score ($T4 - T1$) of the sojourning sample for each broad trait and again show that not all individuals changed in the same direction or to the same extent.. The dashed line represents the mid-point (i.e., no change)

Figure 6: Broad Trait Personality Trait Change Histograms ($T4 - T1$)



5.3.3.1 Qualitative analysis

During the focus group discussions, participants were asked whether they had perceived any changes in their personalities during their year abroad. Each theme will be discussed in turn and then placed within the context of the associated broad traits, with interpretation supported by quotations. This triangulation of results provides a deeper level of understanding of the phenomena under study (Ellis, 1994) and captures data insights not readily available through means of quantitative instruments (Winke, 2017). Nonetheless, the analysis and discussion below are limited by the small sample, and findings lack generalisability due to the convenience sampling technique employed.

Increase in self-confidence

Participants noted that throughout the year abroad, they developed in self-confidence and that this growth facilitated their personal growth in two ways. Firstly, individuals felt they were more confident in their ability to deal with unknown or difficult situations, and secondly, individuals had a stronger desire to try new things and meet new friends.

Ability to deal with the unknown

The ability to deal with unknown or difficult situations can be most closely associated with a decrease in neuroticism, as observed in the quantitative findings. Individuals in all three focus groups spoke of how they felt better equipped emotionally to deal with unfamiliar situations because the year abroad continually pushed individuals out of their comfort zone. For example, participant 8 highlighted how interactions in a non-mutually intelligible language made them appreciate their ability to overcome challenges and difficulties.

“There are also times on the year abroad where you may be stressed or anxious, but you can’t afford not to do something—for example, trying to get someone’s help at a train station when neither of you speak the same language. It’s not something I feel comfortable doing, but the year abroad pushes you into those sorts of scenarios. When you return home, stressful or anxious situations don’t appear so stressful or worrying.” (participant 8)

Participant 3 now appreciates how enjoyable entering new and unfamiliar situations can be. Consequently, participant 2 has fewer worries regarding their future endeavours.

“I feel less worried about the future, I have done something out of my comfort zone, and I enjoyed it. I am looking forward to getting a job, but I’m not as worried about all this as much before.” (participant 3)

Participant 6 highlighted how living with strangers while abroad meant they returned home less anxious about living with new people at their host university. For many returners, finding accommodation can be difficult, for friends met in earlier years have now graduated. Sojourners often have to live with flatmates who are younger and already have substantial social networks in place, and it can be difficult to integrate into these. Nonetheless, the year abroad can better prepare learners to overcome these issues.

“For me, I have become more open to change and not so stressed when placed in a new or unfamiliar environment. This year, I’m living with different people again, and it is now a lot easier that I have gone to France and lived with a bunch of people I didn’t know.” (participant 6)

Desire to try new things and to interact with new people

Most closely aligned with trait openness and trait extraversion, participants, in all three focus groups expressed an increased motivation to seek out new experiences and meet new people. As noted by participant 6 below, this behaviour is in reaction to the need to make new friends and facilitate acculturation and integration into the host community. This finding supports past literature in viewing sojourning as an intervention which can promote learner engagement in a host community, open-mindedness and *broadens horizons* (Mikulec, Jasper, & Cline, 2019; Montgomery & Arensdorf, 2012; Smith & Yang, 2017; Tracy-Ventura et al., 2016).

“I have definitely noticed that I now seek out new opportunities and much more open to going to new societies and inviting people over, even if I have only met them once because when you are abroad, that is how you make friends!” (participant 6)

“I think talkativeness. I find it is a lot easier to inject myself into a conversation.” (participant 5)

“I think my curiosity increased, although I would say I am curious person anyway, but I find myself a lot more open to new experiences.” (participant 11)

Empathy/Compassion

Two participants gave comments associated with empathy and compassion towards other people, with these characteristics closely associated with trait agreeableness. Participant 9 noted how they had become much more appreciative of life as a result of living in a different culture. Here, seeing other models of behaviours in conditions much worse than those found in the UK, made them reflect on their behaviours and made them see themselves differently. Their comment demonstrates an ability “to take both an emic and ethic approach to understanding *their own behaviour and personality*” (Tracy-Ventura et al., 2016, p. 120).

“I think my change [in personality] is through mixing with people who live in worst conditions to me through no fault of their own learn to live with those conditions and adapt. I now consider myself a diva and I have come back home with the perspective that actually I’m really lucky and life isn’t that difficult.” (participant 9)

Participant 11 also felt they had become more aligned with the emotions of other people.

“I felt myself becoming more attuned with people’s emotions and becoming more empathetic.” (participant 11)

While individuals may have become more tolerant of others in foreign cultures, they appeared less tolerant of their fellow learners. Here, the comments below highlight a shift in attitudes

regarding what constitutes as a negative experience. They describe a sense of frustration and have become less empathetic towards the issues faced by non-sojourners. These points reflect the nature of negative events experienced on the year abroad and may demonstrate a level of maturity in behaviours and attitudes not yet present in non-sojourners. It should be noted that this finding does not reflect an intolerance towards other cultures or countries, but rather a reflection that the threshold for what is a stressful event has changed for these individuals, and more of a reflection of increased resilience to perceived problems.

“I think what has caused me frustration has been that I found myself to be a little less tolerant of other people. Hearing people complain, for example, that they don’t have anyone they know in their lecture and that they will be lonely is really frustrating because that isn’t really a problem after facing what I have faced on a year abroad.”
(participant 1)

“I think the things that people find upsetting are now less worthy in my eyes of actually being upset over compared to what I experienced on my year abroad.” (participant 1)

“I think I’m a little dismissive of people’s problems because the threshold of what is a problem has now changed.”
(participant 2)

Autonomy

Two individuals spoke of becoming more independent as a result of the experience abroad. Becoming more autonomous is a widely cited behavioural change stimulated by an experience abroad (Mikulec et al., 2019) and is commonly associated with a higher level of psychological maturity (Weatherley, 1964; Wood et al., 2018). As highlighted by participant 2 and 5, this change may be facilitated by sojourning because familiar support networks are no longer in place, and as such, learners begin to learn that they are no longer dependent on anyone else.

“I think definitely the ability to decide spontaneously on your own that you are going to do this, or you are going to do that because when on a year abroad you have no one telling you what to do and to just get out there and be confident.” (participant 2)

“it [the year abroad] made me more confidence to do things and also live independently because I had to do things on my own.” (participant 5)

Work-ethic

Participant 3 mentioned the difficulties faced in maintaining focus and drive, characteristics associated with conscientiousness. The quantitative data found the average sojourner at university to become less conscientious over time with this finding supporting that of participant 3, who describes rarely being busy while abroad and appreciates the structure offered by their home institution. Early research by Carsello and Creaser (1976) reported how individuals returned home with poorer study habits, while participants in Forsey, Broomhall and Davis (2012) study noted that they rarely attended class, viewing the sojourn experience as a “break from serious study.” Overall, however, the topic of work-ethic has to date, rarely been investigated. It should be noted that participant 3 was a student, and it was unfortunate not to gain the insights of an individual who was employed, as employment status may have impacted perceived change in conscientiousness, a trait associated with work-ethic (e.g., Specht et al., 2011) as observed in section 5.3.4.2.

“I feel a lot more focused having come back. My year abroad didn’t really have a focus. I studied, but I didn’t have exams. I worked but didn’t have a lot to do to, and I had no focus. Upon returning, I have a lot more focus and appreciate being busy, and my time-management is much better.” (participant 3)

To grow or not to grow?

Not everybody perceived personal growth. When asked to expand on why they believed this to be, both participants 7 and 12 indicated that they felt this was because their host countries were culturally not dissimilar to the UK. While countries such as Germany and Japan are considered *abroad*, given their westernised values, the experiences and opportunities available to these learners may not be dramatically different. This finding supports the view that the year abroad may not be as culturally immersive as it once was (Coleman, 2015; Jackson, 2018) and further supports the quantitative findings regarding individual differences in trait change. It should be noted, however, that both these individuals were students at an international university with contact hours in English. Perhaps, these individuals were not afforded the opportunities to become immersed in the culture to the same extent as those who worked or taught.

“I wouldn’t say I quite found that because Germany isn’t too different from here and as such didn’t really perceive any great changes.” (participant 7)

“I don’t know because Japan is very similar to Britain with regards to their morals and outlook on life, so it is quite similar but in a Japanese way and so I don’t know if I have changed.” (participant 12)

Participant 10 explained this lack of change was because they had already been through a similar experience and they already felt prepared emotionally for the challenges of the year abroad. As noted by Tracy-Ventura et al. (2016), many stakeholders perceive the year abroad to be the first culturally immersive experience for *all* who undertake it, when actually for some, it is not (e.g., heritage language learners), and this deserves further discussion in the literature.

“Before coming to university, I worked and lived abroad, and I think I changed as a person most then, which for a lot of people the year abroad is the first time they have been abroad. So, I feel I have done a lot of the emotional changes already and felt prepared when going into the year abroad.” (participant 10)

5.3.3.2 Summary of findings

Section 5.3.3 has evaluated whether a) broad traits change significantly over time and b) whether significant individual differences exist in the extent to which broad trait personality change.

The findings have shown sojourners to return home significantly more open, agreeable, and emotionally stable, while non-sojourners became significantly more extraverted over time. Due to the lack of comparison group, these main effects of time (apart from agreeableness) cannot be causally attributed to the learning context. Moreover, the findings identified substantial individual differences in all broad trait trajectories, which warrant further investigation. The qualitative data indicated that not all sojourners perceived growth, and in cases where growth was perceived, it could not always be considered socially desirable.

5.3.4 Predictors of individual differences in broad trait trajectories

The above findings have demonstrated substantial individual differences in trait change trajectories across time. Understanding why these differences exist warrant further attention, with analysis guided by the following research question and examined sojourner data only:

RQ4: What are the predictors of broad trait change?

This research question was answered through a series of multiple regression analyses. For each regression, a change score (post-test – pre-test) variable served as the dependent variable, while a series of predictors served as the independent variables. A change-score approach was undertaken because the analysis was interested in understanding why some individuals *changed* more than others in relation to their broad personality traits. Such an approach is similar to that undertaken by Cubillos and Ilvento (2013) and Hessel and Vanderplank (2018), both of whom were interested in ascertaining predictors of individual differences in language gain. Predictors were divided across three types of variable. The first was associated with the *person* and comprised of the baseline score of each of the five broad traits. These factors can be considered independent of the year abroad experience, for they were collected prior to departure for the majority of sojourners. *Environmental* factors were also considered and refer to factors which relate to how sojourners perceived and interpreted their immediate environment throughout the academic year. Lastly, factors associated with the *sojourn programme* were considered.

Where possible, all valid cases were used. Out of the 110 sojourners, ten failed to complete the trait measure either at T1 or T4, and a change score for these individuals could not be calculated. They are as such dropped from this analysis. Given the sample size, it was decided that up to five predictors could be entered into the regression model simultaneously. As most learners departed in September ($n = 48$) and returned in June ($n = 53$), the analysis below concerns data collected during these months (i.e., ten months).

Given the sample size, it was decided that up-to five predictors could be entered into the regression model simultaneously. As most learners departed in September ($n = 48$) and returned in June ($n = 53$), the analysis below concerns data collected during these months.

Prior to conducting the hierarchical regression for each of the broad traits, a series of smaller multiple regression analyses were undertaken. Table 42 outlines all factors which were considered in the analysis, divided across the three areas previously noted (e.g., *person*). Each trait change score was regressed on a series of variables while holding the corresponding trait baseline score constant. Factors found to be significant at this stage were carried forward to the

hierarchical regression model. Had the sample size been larger, it would have been preferable to first enter all *person*-related factors, followed by *environmental* and *programme-specific* factors together.

Table 42: Testing Potential Covariates of Broad Trait Change from T1 to T4 (Sojourners)

Variable	t-statistic and sig. (each dependent variable (i.e., openness) is a change score)				
	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Baseline broad personality traits					
Openness	-	t (99) = -1.260, p = .21	t (100) = .178, p = .85	t (97) = .796, p = .42	t (100) = -1.898, p = .06
Conscientiousness	t (100) = 1.446, p = .15	-	t (100) = 2.565, p = .01*	t (97) = 2.236, p = .02*	t (100) = -1.857, p = .06
Extraversion	t (100) = -2.034, p = .04*	t (99) = -.728, p = .46	-	t (97) = -1.006, p = .31	t (100) = .165, p = .86
Agreeableness	t (100) = 2.178, p = .03*	t (99) = -.361, p = .71	t (100) = .274, p = .78	-	t (100) = -2.055, p = .04*
Neuroticism	t (100) = -.111, p = .91	t (99) = -.991, p = .32	t (100) = -.813, p = .41	t (97) = -1.750, p = .08	-
Environmental/perceived behaviours					
Loneliness	t (100) = -2.831, p = .006*	t (99) = -2.047, p = .04*	t (100) = -3.100, p = .002*	t (97) = -1.964, p = .05	t (100) = 3.111, p = .002
Friendship	t (100) = 1.771, p = .08	t (99) = 1.039, p = .30	t (100) = 1.813, p = .07	t (97) = .906, p = .36	t (100) = -.740, p = .46
Belongingness	t (100) = 1.210, p = .22	t (99) = 2.111, p = .03*	t (100) = 2.434, p = .01*	t (97) = .164, p = .87	t (100) = -.937, p = .35
Negative event	t (100) = -.131, p = .89	t (99) = -1.740, p = .08	t (100) = -.433, p = .66	t (97) = -2.113, p = .03*	t (100) = 2.726, p = .007*
Participation in extra-cur. activities	t (100) = 1.938, p = .05	t (99) = 2.060, p = .04*	t (100) = 2.537, p = .01*	t (97) = 1.712, p = .09	t (100) = -2.667, p = .009*
Travelled	t (100) = .033, p = .97	t (99) = -.503, p = .96	t (100) = .215, p = .83	t (97) = -.463, p = .64	t (100) = -.279, p = .78
Program characteristics					
Previous Experience	t (100) = .219, p = .82	t (99) = -.531, p = .59	t (100) = 1.546, p = .12	t (97) = .746, p = .45	t (100) = .191, p = .84
Language Student	t (100) = -.092, p = .92	t (99) = .783, p = .43	t (100) = .017, p = .98	t (97) = -.227, p = .82	t (100) = -1.008, p = .31
Stayed in two countries	t (100) = .718, p = .47	t (99) = 1.027, p = .30	t (100) = -.437, p = .66	t (97) = -1.135, p = .25	t (100) = 1.040, p = .30
Stayed in Europe	t (100) = .025, p = .98	t (99) = .918, p = .36	t (100) = -1.018, p = .31	t (97) = .265, p = .79	t (100) = .039, p = .96
Comp/Vol	t (100) = .215, p = .83	t (99) = 1.119, p = .26	t (100) = .006, p = .99	t (97) = -.232, p = .81	t (100) = -1.061, p = .29
Sojourner Role	t (100) = .870, p = .38	t (99) = 2.239, p = .02*	t (100) = 1.061, p = .29	t (97) = .207, p = .83	t (100) = .075, p = .94
Length of Stay	t (100) = .067, p = .94	t (99) = .021, p = .98	t (100) = .339, p = .73	t (97) = -1.493, p = .13	t (100) = 1.026, p = .30

Key: *p = <.05

For each broad trait, the regression model was built by first entering the learner’s baseline corresponding broad trait score, so that the variance explained by this score was controlled for. Next, *person* factors were added, followed by *environmental* and lastly *programme-specific* factors. In doing so, it was possible to ascertain the amount of variance accounted for by each factor.

5.3.4.1 Change in trait openness

The preliminary regression analysis (Table 42) indicated that baseline extraversion (partial $r = -.19$), baseline agreeableness (partial $r = .21$) and loneliness (partial $r = -.27$) were all significantly associated with openness change after controlling for baseline openness. While participation in extra-curricular activities did not reach significance ($p = .05$) when independently regressed, it was retained in the final model given its closeness to the boundary of significance ($p = .05$).

Table 43 presents the output. Regarding assumptions, linearity was achieved, as assessed by partial regression plots. A Durbin-Watson statistic of 1.982 indicated independence of residuals. Homoscedasticity was also achieved, and there was no evidence of multicollinearity as assessed by no tolerance value lower than 0.1. There were no studentised deleted residuals greater than ± 3 standard deviations. The assumption of normality was met, as assessed by Q-Q Plot.

Table 43: Multiple Regression of Change in Openness on the Identified Covariates (Sojourners)

Factors	Unstan. Coefficients		Standardised coefficients				zero-order	partial
	B	SE	Beta	T	Sig.			
Model 1								
Constant	.689 (.129, 1.248)	.282		2.442	.016			
Openness T1	-.154 (-.300, -.008)	.074	-.207	-2.093	.039	-.207	-.207	
Model 2								
Constant	.312 (-.403, 1.027)	.360		.867	.388			
Openness T1	-.149 (-.289, -.009)	.071	-.200	-2.106	.038	-.207	-.210	
Extraversion T1	-.118 (-.204, -.032)	.043	-.266	-2.722	.008	-.223	-.268	
Agreeableness T1	.195 (.059, .332)	.069	.275	2.834	.006	.190	.278	
Model 3								
Constant	.596 (-.107, 1.299)	.354		1.683	.096			
Openness T1	-.161 (-.294, -.029)	.067	-.217	-2.412	.018	-.207	-.241	
Extraversion T1	-.131 (-.213, -.049)	.041	-.294	-3.172	.002	-.223	-.311	
Agreeableness T1	.183 (.054, .313)	.065	.258	2.806	.006	.190	.278	
Loneliness	-.373 (-.636, -.110)	.133	-.251	-2.813	.006	-.269	-.279	
Participation	.242 (.027, .458)	.108	.201	2.234	.028	.166	.225	

Model 1: $R^2 = .043$, Adjusted $R^2 = .033$, $F(1, 98) = 4.382$, $p = .03$; Model 2: $R^2 = .153$, Adjusted $R^2 = .126$, change in $R^2 = .110$, change in $F(2, 96) = 6.233$, $p = .003$; Model 3: $R^2 = .260$, Adjusted $R^2 = .221$, change in $R^2 = .107$, change in $F(2, 94) = 6.827$, $p = .002$

As shown in Table 43, baseline openness indicated 4.3% of the variation in students' overall openness change abroad (Model 1, $p = .03$). The model's capacity to predict openness change was significantly improved by adding both *person* (Model 2, $p = .003$) and *environmental* (Model 3, $p = .002$) related factors. Model 3 was chosen as the final model as it explained 26.0% of the variation in sojourners' overall openness change across the year abroad.

Attention is now be given to those factors which made a significant contribution in the final model in accounting for openness change in the sojourn sample.

Baseline openness

Baseline openness score was shown to be a significant predictor of openness change ($t(100) = -2.412, p = .018$), accounting for 4.3% of variation in individuals' change scores. The negative partial correlation score (partial $r = -.24$) indicated that sojourners who scored low in openness at the beginning of the year abroad tended to show more substantial change than those who initially scored high. This association is to be expected and is possibly reflective of the phenomenon known as the "natural regression towards the mean" (Allison, 1990). As such, it is assumed that this result is, to at least some extent, a result of a statistical artefact.

Baseline extraversion

Baseline extraversion demonstrated the strongest association with openness change ($t(100) = -3.172, p = .018$, partial $r = -.31$), with more introverted learners at baseline displaying greater change in openness. This finding is not conducive to immediate explanation. At baseline, the correlational analysis indicated only a weak relationship between openness and extraversion ($p = .31, r = .08$), suggesting those who were higher in openness at baseline, were not necessarily higher in extraversion. The question, therefore, arises as to why more introverted individuals are more likely to become more open as a result of the year abroad experience compared to individuals who perceived themselves higher in extraversion at baseline? In the at-home context, while the relationship between the two variables was negative, it failed to reach significance when modelled independently ($t(65) = -.585, p = .56$, partial $r = -.07$), indicating it may be something about the learning context which facilitates this change to a greater extent. From a theoretical perspective, extraverts tend to seek out situations which fulfil their need for affiliation and exhibition, whereas introverts, will tend to seek out small groups or be happy to do things alone (Oishi, Talhelm, & Lee, 2015). Previous literature (e.g., Ozer, 2015; Severiens & Schmidt, 2009) has long discussed the difficulties sojourners face when trying to integrate into a new host

community. As such, the year abroad environment may not be conducive to large group activities, and instead, individuals who are comfortable to explore new and unfamiliar environments on their own or in small groups may benefit more from the experiences afforded to them on the year abroad. This is one explanation posited and warrants further investigation.

Baseline agreeableness

Baseline agreeableness was also found to predict openness change significantly ($t(100) = 2.806, p = .006$, partial $r = .27$), suggesting those more agreeable at baseline tended to demonstrate larger changes in openness than those less agreeable. As noted by Schumann (1978), agreeable learners are more likely to acculturate at a faster rate and as such, be afforded new cultural opportunities not readily available to *outsiders*. Such opportunities can prove culturally stimulating and increase learners' desire to seek and appreciate new and novel experiences. Moreover, agreeableness has been linked with cultural intelligence. Due to their high level of interpersonal competencies, agreeable learners are more likely to learn from others who are culturally different and be stimulated in doing so (Li, Mobley, & Kelly, 2016).

Loneliness

Loneliness was significantly and negatively associated with openness change ($p = .006$, partial $r = -.27$). Few studies have directly explored the relationship, and none have done so in a study abroad context, to the best of my knowledge. In a recently published meta-analysis, Buecker, Maes, Denissen and Luhmann (2020) found openness to show a weak, significant association with loneliness ($r = -.10$), but which, when controlling for the remaining four broad traits were non-significant (95% CI [-.038, .039]). In this current study, those at-home demonstrated not only a non-significant association between openness change and loneliness ($t(65) = 1.007, p = .31$) but an inverse relationship also. It may therefore be posited that the specificities of the sojourning context exacerbate the impact loneliness has on an individuals' openness change. This negative association may be indicative of lonely individuals being afforded lessened opportunities to seek out new experiences, together with having lower motivation to seek out opportunities.

Participation in extracurricular activities

The addition of students' extracurricular activities made a significant contribution to the overall predictive capacity of the model ($t(100) = 2.234, p = .02$). The positive association (partial $r = .22$) indicated those who more frequently partook in extra-curricular activities tended to become more open than those who chose not to. This finding is consistent with the Correspondive

Principle of personality development (see Roberts et al., 2006b) which posits “a reciprocal relationship between personality traits and life experiences” (Harms, 2019, p. 1). Past research has suggested those high in trait openness will more enthusiastically be involved in extracurricular activities (Shiner & DeYoung, 2013; Stephan, Boiché, Canada, & Terracciano, 2014). Going forth, therefore, these experiences further stimulate positive change in openness. Moreover, it may be posited that participation in such activities can foster intercultural curiosity, aid integration and again afford learners new culturally stimulating experiences which otherwise would not be immediately available to them. The finding that participation in extracurricular activities did not predict openness change in those at-home ($t(65) = .767, p = .44, \text{partial } r = .08$), further suggests that there may be something specific about the sojourning context which allows extra-curricular participation to facilitate openness change.

5.3.4.2 *Change in trait conscientiousness*

The covariate table (Table 42) identified several significant predictors of conscientiousness change. When controlling for baseline conscientiousness score, loneliness (partial $r = -.20$), belongingness (partial $r = .21$) and club participation (partial $r = .20$) were all significant predictors of change. Moreover, sojourner role was also found to be significant ($t(99) = 2.259, p = .02$). As this variable was dichotomously coded, the finding suggested that those in employment returned home significantly more conscientious than those who studied.

Table 44 presents the hierarchical model, with potential covariates added at each level. Regarding assumptions, there was linearity, as assessed by partial regression plots and a plot of studentised residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.885. There was homoscedasticity, as assessed by visual inspection of a plot of studentised residuals versus unstandardised predicted values. There was no evidence of multicollinearity as assessed by an average VIF value of 1.041 and a tolerance value of .960. There was one outlier as with a studentised deleted residual greater than ± 3 standard deviations, and this case was dropped from the analysis. The assumption of normality was met, as assessed by Q-Q Plot.

Table 44: Multiple Regression of Change in Conscientiousness on the Identified Covariates (Sojourners)

Factors	Unstan. Coefficients		Standardised coefficients				
	B	SE	Beta	T	Sig.	zero-order	partial
Model 1							
Constant	.042 (-.436, .519)	.240		.173	.863		
Conscientiousness T1	-.012 (-.139, .115)	.064	-.018	-.182	.856	-.018	-.018
Model 2							
Constant	.015 (-.584, .614)	.302		.050	.960		
Conscientiousness T1	-.055 (-.180, .070)	.063	-.087	-.868	.387	-.018	-.089
Loneliness	-.249 (-.520, .022)	.137	-.182	-1.823	.071	-.196	-.185
Belongingness	.311 (-.025, .647)	.169	.180	1.839	.069	.209	.186
Participation	.202 (-.013, .418)	.108	.183	1.865	.065	.203	.189
Model 3							
Constant	-.035 (-.623, .553)	.296		-.118	.907		
Conscientiousness T1	-.045 (-.168, .078)	.062	-.071	-.726	.470	-.018	-.075
Loneliness	-.273 (-.540, -.007)	.134	-.200	-2.036	.045	-.196	-.207
Belongingness	.264 (-.068, .596)	.167	.153	1.578	.118	.209	.161
Participation	.200 (-.011, .410)	.106	.181	1.879	.063	.203	.191
Sojourner Role	.155 (.017, .292)	.069	.215	2.231	.028	.224	.225

Model 1: $R^2 < .001$, Adjusted $R^2 < .001$, $F(1, 79) = .033$, $p = .856$; Model 2: $R^2 = .112$, Adjusted $R^2 = .074$, change in $R^2 = .112$, change in $F(3, 94) = 3.944$, $p = .011$; Model 3: $R^2 = .157$, Adjusted $R^2 = .112$, change in $R^2 = .045$, change in $F(1, 93) = 4.976$, $p = .028$

The output (Table 44) showed that baseline conscientiousness explained little to no variation in sojourners' conscientiousness change (model 1, $p = .85$) suggesting baseline conscientiousness to have little effect on subsequent change. No further *person* variable was found to be significantly related to the outcome variable. The addition of *environmental* factors (model 2, $p = .01$) and *program characteristics* (model 3, $p = .02$) did, however, make a significant contribution to the predictive quality of the model. The final model (model 3) explained a total of 15.7% of the variation in overall sojourner conscientiousness change across ten months.

After controlling for all relevant variables, baseline conscientiousness change ($p = .47$), belongingness ($p = .11$) and participation in extracurricular activities ($p = .06$) were no longer significant predictors of change. This indicated that the variance in the outcome explained by these variables was now in some way explained by the two remaining significant variables (loneliness and sojourner role).

Loneliness

The addition of loneliness made a significant, unique contribution to the prediction of overall conscientiousness change ($t(99) = -2.036$, $p = .04$). Loneliness was negatively associated with

overall conscientiousness change (partial $r = -.20$), indicating that those who perceived themselves as less lonely while abroad, became more conscientious over time. The link between conscientiousness and loneliness is rather tenuous, and there has been no conclusive evidence regarding the directionality of this relationship (Mund & Neyer, 2019; Vanhalst et al., 2012). Theoretically, conscientious individuals are more likely to maintain contact with friends and family while abroad (Buecker et al., 2020) consequently minimising loneliness. As such, the corresponive principle would posit that less conscientious individuals may feel lonelier, which in turn can inhibit conscientious like behaviours further.

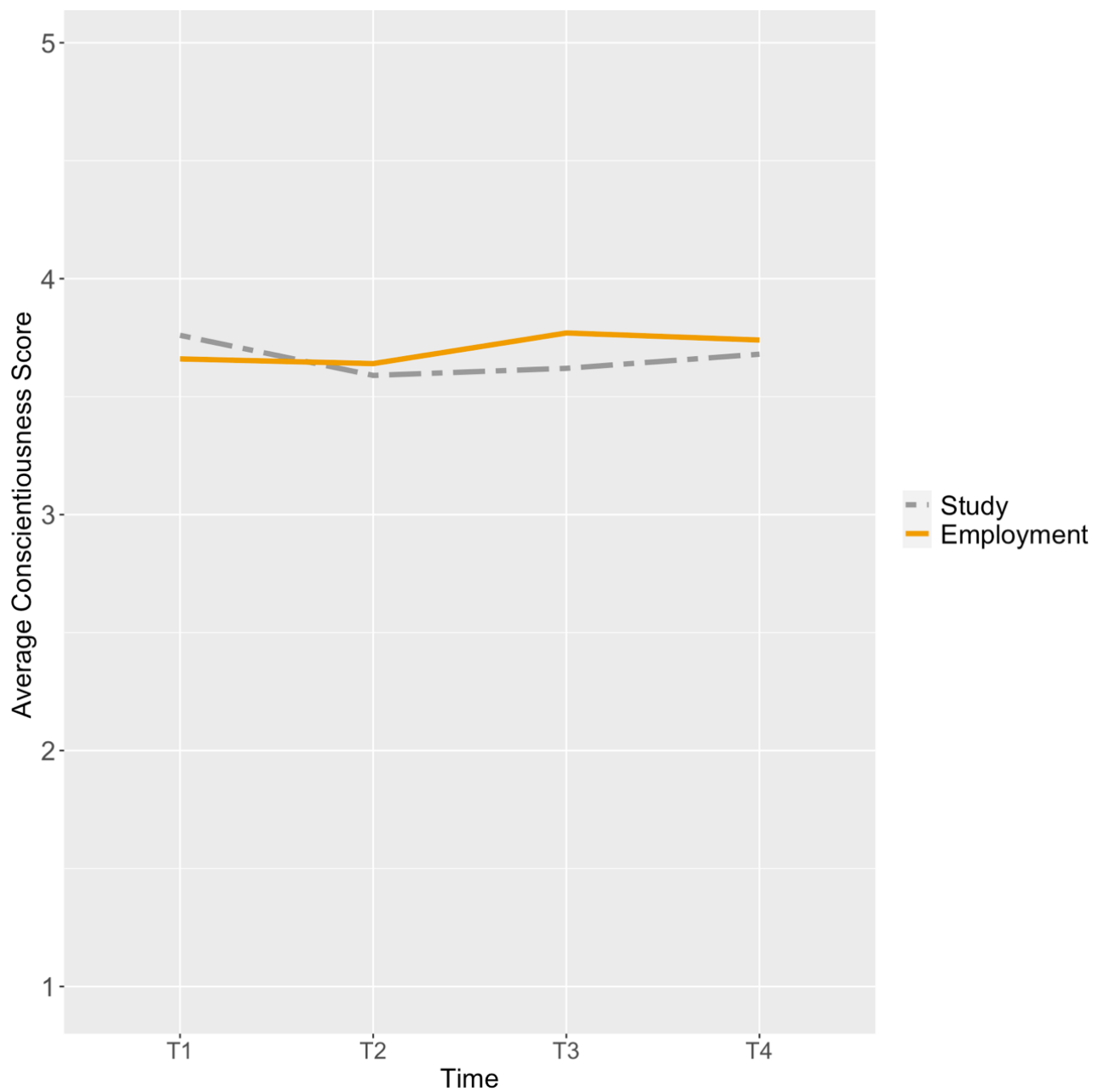
Sojourner role

Sojourner role was found to be significantly associated with overall conscientiousness change ($t(81) = 2.232, p = .011, \text{partial } r = .24$). As those who studied were coded with 0, and those in paid employment with 1, the findings indicate that individuals who studied for the entire year tended to show smaller changes in conscientiousness when their baseline conscientiousness score was equal. Paid employment has typically been associated with increasing conscientiousness (e.g., Specht et al., 2011), yet literature specific to the sojourn experience (Greischel et al., 2016; Niehoff et al., 2017; Zimmermann & Neyer, 2013) has tended to show little change in conscientiousness over time. This finding has puzzled researchers due to the associated acceleration in maturation demonstrated in the other traits.

Further analysis was given to understanding conscientiousness change in the sojourn sample. A two-way mixed ANOVA was undertaken in order to ascertain whether mean conscientiousness scores were significantly higher in those in paid employment over time compared to those who studied. There was a statistically significant interaction between sojourner role and time on conscientiousness score $F(3, 288) = 3.561, p = .015, \text{partial } \eta^2 = 0.36$. Bonferroni corrected applied mean scores indicated that the average student declined by $-.04$ units more in conscientiousness over time compared to those in employment. The simple main effects indicated a significant effect for time for those who studied while abroad (i.e., at a university) ($p = .01$), but not for those in employment ($p = .06$). Here, posthoc testing showed university students significantly declined in conscientiousness ($p = .03$) between T1 ($m = 3.76; sd = .084$) and T2 ($m = 3.59; sd = .097$). There was no significant effect for sojourner role with learners in each type of role perceiving similar conscientiousness scores at each timepoint (T1: $p = .39$; T2: $p = .70$; T3: $p = .23$; T4: $p = .62$). The general decline in conscientiousness for those at university support the findings of Zimmermann and Neyer (2013) and Niehoff et al. (2017), who both

found those who studied tended to become less conscientious over the sojourn period. The findings of this current study have provided an insight into those in paid employment on a year abroad, which to date has not been forthcoming. The findings indicate that paid employment can foster greater conscientiousness change over that of studying while abroad. A significant main effect for *time* may have been found for those in paid employment, had the sample been larger. Figure 7 displays this interaction between time and sojourner role.

Figure 7: Conscientiousness Change Interaction Effect (*Employment Status * Time*) in Sojourner Sample



5.3.4.3 Change in trait extraversion

When exploring potential covariates of extraversion change (Table 42), baseline conscientiousness scores (partial $r = .25$), loneliness (partial $r = .30$), belongingness (partial $r = .24$), and extra-curricular participation (partial $r = .24$), were all found to be significantly associated with change and were therefore carried forward to the final model.

Table 45 presents the results of the hierarchical regression analysis for trait extraversion change. Regarding assumptions, linearity was achieved, together with independence of residuals, as assessed by a Durbin-Watson statistic of 1.860. Homoscedasticity was present. There was no evidence of multicollinearity as assessed by an average VIF value of 1.104 and a tolerance value of .930. There were no studentised deleted residuals greater than ± 3 standard deviations. The assumption of normality was met, as assessed by Q-Q Plot.

As shown in Table 45, baseline extraversion scored explained 19.1% of the variation in the sojourners' overall extraversion change (Model 1, $p = <.001$). The prediction of extraversion change was significantly improved by adding both *person*-related factors (Model 2, $p = .01$) and related *environmental* factors (Model 3, $p = .001$). The final model (model 3) explained 36.7% of the variation in student's overall extraversion change across the year.

Table 45: Multiple Regression of Change in Extraversion on the Identified Covariates (Sojourners)

Factors	Unstan. Coefficients		Standardised coefficients				
	B	SE	Beta	T	Sig.	zero-order	partial
Model 1							
Constant	1.153 (.714, 1.592)	.221		5.217	<.001		
Extraversion T1	-.302 (-.426, -.177)	.063	-.437	-4.813	<.001	-.437	-.437
Model 2							
Constant	.365 (-.379, 1.109)	.375		.973	.333		
Extraversion T1	-.346 (-.471, -.220)	.063	-.500	-5.455	<.001	-.437	-.485
Conscientiousness T1	.253 (.057, .448)	.098	.235	2.565	.012	.101	.252
Model 3							
Constant	.418 (-.451, 1.287)	.438		.954	.342		
Extraversion T1	-.380 (-.500, -.261)	.060	-.551	-6.336	<.001	-.437	-.547
Conscientiousness T1	.189 (.003, .374)	.094	.176	2.018	.046	.101	.204
Loneliness	-.479 (-.866, -.093)	.195	-.208	-2.463	.016	-.252	-.246
Belongingness	.546 (.056, 1.036)	.247	.186	2.211	.029	.131	.222
Participation	.356 (.047, .665)	.156	.190	2.286	.024	.168	.230

Model 1: $R^2 = .191$, Adjusted $R^2 = .183$, $F(1, 98) = 23.161$, $p = <.001$; Model 2: $R^2 = .243$, Adjusted $R^2 = .227$, change in $R^2 = .051$, change in $F(1, 97) = 6.581$, $p = .012$; Model 3: $R^2 = .367$, Adjusted $R^2 = .334$, change in $R^2 = .125$, change in $F(3, 94) = 6.176$, $p = .001$

Baseline extraversion

The sojourners' baseline extraversion score served as the strongest predictor within Model 3 ($t(100) = -6.336, p = <.001$) and explained 19.7% of the variation in sojourners' overall extraversion change. The negative association (partial $r = -.54$) indicated those who began the year abroad more introverted, tended to become more extroverted during the year abroad. Similar to that of openness, this finding may be caused by a statistical artefact (i.e., natural regression to the mean). Nonetheless, from a theoretical perspective, extraversion is an important trait in facilitating integration into the host community through the establishment of social networks (Ward, Bochner, & Furnham, 2001). Consequently, those learners more introverted at baseline may have made both conscious and sub-conscious decision to alter their thoughts, feelings and behaviour over time in order to facilitate integration, as demonstrated by becoming more extraverted.

Baseline conscientiousness

Baseline conscientiousness was also positively associated with extraversion change ($t(100) = 2.018, p = .04$, partial $r = .20$). A direct explanation between baseline conscientiousness and extraversion change is not forthcoming in the literature, and it may be possible that a number of variables are mediating this relationship. For example, conscientiousness and extraversion have both been strongly linked with subjective well-being (Fayard, Roberts, Robins, & Watson, 2012).

Loneliness

A learner's average loneliness score was found to be significantly and negatively associated ($t(100) = -2.463, p = .01$, partial $r = .24$). with overall extraversion change, indicating those who felt lonelier on a year abroad, tended to show a smaller gain. While loneliness is well explored in a study abroad context (e.g., Hunley, 2010; Wiseman, 1997), its influence on personality change, has, to the best of my knowledge, not yet been ascertained. Nonetheless, the relationship found is consistent with the view that loneliness limits the opportunities individuals have to seek out social interaction and limits the availability of support networks, both of which would inhibit positive extraversion change (Fleeson & Gallagher, 2009; Selfhout et al., 2010). It may be further posited that due to entering into a new host community, the impact of loneliness on extraversion may be greater than if remaining at-home, where cultural and linguistic differences are not present to the same extent, and the ability to gain a social position perhaps easier.

Belongingness

Belongingness was positively and significantly associated with extraversion change ($p = .02$, partial $r = .22$). This association is reflective of the relationship between social connectedness and extraversion change, whereby those who feel socially connected, will demonstrate a stronger desire to seek out social interaction (Hanna, Tefertiller, & Cota, 2015).

Participation in extracurricular activities

Adding information on extracurricular participation made a significant unique contribution to the prediction of extraversion change ($t(81) = 2.40, p = .01$). The two variables were positively associated with each other (partial $r = .27$) suggesting that those who took part in extracurricular activities tended to make larger changes in extraversion across the year. Literature has long associated higher levels of extraversion with a stronger preference for activities requiring socialization (e.g., Asendorpf & Wilpers, 1998; Furnham, Zhang, & Chamorro-Premuzic, 2006; Oerlemans, Bakker, & Veenhoven, 2011), but to the best of my knowledge, no study has investigated a direct link between participation in extracurricular activities and extraversion change. Within the sojourning context, it may be expected that participation provides an opportunity to meet new people and form new social circles. Theoretically, the corresposive principle posits that these individuals will behave extroverted in the first place (i.e., when joining a club) and that through participating in clubs, extroverted behaviour will be stimulated.

5.3.4.4 Change in trait agreeableness

The preliminary regression analysis (Table 42) showed baseline conscientiousness (partial $r = .22$), Loneliness (partial $r = -.19$) and the onset of negative events (partial $r = -.21$) to all be significantly associated with changes in trait agreeableness.

Table 46 present the hierarchical model as built following the procedure earlier described. Regarding assumptions, there was linearity, as assessed by partial regression plots and a plot of studentised residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.729, falling within the suggested range of 1-3. There was homoscedasticity, as assessed by visual inspection of a plot of studentised residuals versus unstandardised predicted values. There was no evidence of multicollinearity as assessed by no tolerance value under 0.100. There were three outliers with a studentised deleted residual greater than ± 3 standard deviations. The assumption of normality was met, as assessed by Q-Q Plot.

Table 46: Multiple Regression of Change in Agreeableness on the Identified Covariates (Sojourners)

Factors	Unstan. Coefficients		Standardised coefficients				
	B	SE	Beta	T	Sig.	zero-order	partial
Model 1							
Constant	1.332 (.770, 1.893)	.283		4.709	<.001		
Agreeableness T1	-.309 (-.452, -.166)	.072	-.403	-4.297	<.001	-.403	-.403
Model 2							
Constant	.883 (.204, 1.562)	.342		2.582	.011		
Agreeableness T1	-.339 (-.482, -.197)	.072	-.443	-4.727	<.001	-.403	-.435
Conscientiousness T1	.152 (.017, .288)	.068	.209	2.236	.028	.126	.225
Model 3							
Constant	1.196 (.463, 1.929)	.369		3.241	.002		
Agreeableness T1	-.335 (-.476, -.195)	.071	-.437	-4.738	<.001	-.403	-.443
Conscientiousness T1	.126 (-.010, .262)	.068	.173	1.837	.069	.126	.188
Loneliness	-.173 (-.472, .127)	.151	-.110	-1.145	.255	-.159	-.119
Negative Event	-.307 (-.698, .084)	.197	-.147	-1.561	.122	-.207	-.161

Model 1: $R^2 = .163$, Adjusted $R^2 = .154$, $F(1, 95) = 18.461$, $p < .001$; Model 2: $R^2 = .205$, Adjusted $R^2 = .188$, change in $R^2 = .042$, change in $F(1, 94) = 5.000$, $p = .02$; Model 3: $R^2 = .246$, Adjusted $R^2 = .213$, change in $R^2 = .041$, change in $F(2, 92) = 2.511$, $p = .08$

As shown in Table 46, baseline agreeableness accounted for 16.3% of the variation in the agreeableness change (Model 1, $p < .001$). The addition of *person*-related factors significantly improved the model's predictability (Model 2, $p = .02$), while the addition of *Environmental* factors did not significantly improve the predictability of the regression model (model 3, $p = .08$).

Baseline agreeableness

Baseline agreeableness score was the strongest predictor within the set ($t(81) = -6.00$, $p < .001$), and explained 16.3% of the variation in the outcome variable (agreeableness change). Again, given the negative partial correlation (partial $r = -.44$), it can be said that those with initial lower agreeableness score tended to show more substantial change than those who initially scored higher in agreeableness. While this finding could be put down to statistical artefact (i.e., natural regression towards the mean), this conclusion would appear too simplistic. One possible alternative explanation is that in order to integrate successfully into a new community, one must show stronger trait agreeable responses, such as trustworthiness and friendliness. Chan and Sy (2016) found that out of the traits of agreeableness, openness and conscientiousness; agreeableness showed the strongest relationship with intercultural communication, while Peifer and Yangchen (2017) found agreeableness to be a significant predictor of cultural-emotional intelligence. For non-sojourners, on the other hand, they are already integrated into the environment and have already established strong support networks.

5.3.4.5 Change in trait neuroticism

Table 42 identified five significant covariates of neuroticism change after controlling for baseline neuroticism. These were baseline agreeableness (partial $r = -.20$), loneliness (partial $r = .30$), negative event (partial $r = .26$), and extra-curricular activities (partial $r = -.26$).

Regarding assumptions of regression model found in Table 47, linearity was achieved as assessed by partial regression plots. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.509. There was homoscedasticity, as assessed by visual inspection of a plot of studentised residuals versus unstandardised predicted values. There was no evidence of multicollinearity as assessed by an average VIF value of 1.072 and a tolerance value of .934. There were no studentised deleted residuals greater than ± 3 standard deviations. The assumption of normality was met, as assessed by Q-Q Plot.

Table 47: Multiple Regression of Change in Neuroticism on the Identified Covariates (Sojourners)

Factors	Unstan. Coefficients		Standardised coefficients				
	B	SE	Beta	T	Sig.	zero-order	partial
Model 1							
Constant	.936 (.369, 1.503)	.286		3.277	.001		
Neuroticism T1	-.357 (-.536, -.178)	.090	-.371	-3.949	.000	-.371	-.371
Model 2							
Constant	2.080 (.842, 3.318)	.624		3.336	.001		
Neuroticism T1	-.388 (-.567, -.209)	.090	-.402	-4.298	.000	-.371	-.400
Agreeableness T1	-.270 (-.531, -.009)	.131	-.192	-2.055	.043	-.126	-.204
Model 3							
Constant	1.735 (.573, 2.897)	.585		2.965	.004		
Neuroticism T1	-.465 (-.632, -.297)	.084	-.482	-5.507	.000	-.371	-.494
Agreeableness T1	-.254 (-.491, -.016)	.120	-.181	-2.115	.037	-.126	-.213
Loneliness	.618 (.096, 1.140)	.263	.210	2.353	.021	.199	.236
Negative Event	.797 (.127, 1.468)	.338	.209	2.362	.020	.179	.237
Participation	-.550 (-.950, -.150)	.201	-.231	-2.733	.008	-.238	-.271

Model 1: $R^2 = .137$, Adjusted $R^2 = .128$, $F(1, 98) = 15.595$, $p < .001$; Model 2: $R^2 = .173$, Adjusted $R^2 = .156$, change in $R^2 = .036$, change in $F(1, 97) = 4.224$, $p = .04$; Model 3: $R^2 = .336$, Adjusted $R^2 = .300$, change in $R^2 = .162$, change in $F(3, 94) = 7.655$, $p < .001$

The output of Table 47 indicated that baseline neuroticism explained 13.7% of the variation in sojourners' neuroticism change (Model 1, $p < .001$). The addition of *person* factors (Model 2, $p = .04$) significantly improved the predictive capacity of the model, whereas the addition of *environmental* factors had a non-significant effect on predictability (Model 3, $p < .001$). Model 3 was chosen as the final model as it explained 33.6% of variation in neuroticism scores.

Baseline neuroticism

Baseline neuroticism was the strongest predictor within Model 3 ($t(81) = -6.42, p = <.001$), displaying a negative relationship with overall neuroticism change (partial $r = -.60$). This suggests that individuals with higher baseline Neuroticism scores tended to become more emotionally stable and in line with the previous explanation, maybe as a result of statistical artefact.

Baseline agreeableness

A significant inverse relationship was found between the students' baseline agreeableness score and neuroticism change ($p = .02$, partial $r = .27$), indicating that the higher a sojourner scored in agreeableness, the more likely it is that they will become less neurotic over time. This finding supports previous literature (McCrae & Costa, 1990; Shi et al., 2018) which has investigated the relationship between the two variables. Moreover, within a sojourning context, trait-related behaviours associated with agreeableness will likely aid the integration process. Consequently, it may be anticipated that those who are more agreeable at baseline adapt and integrate faster into the new host community, which has a positive effect on reducing neuroticism over time.

Loneliness

A significant and positive association was found between loneliness and neuroticism change ($t(100) = 2.353, p = .02$, partial $r = .23$), indicating lonelier individuals tended to become more neurotic over time. This finding is in line with previous literature on the subject matter (e.g., Hensley et al., 2012; Vanhalst et al., 2012), but the first to situate itself within a sojourning context, to the best of my knowledge. This relationship is consistent with the view that scoring high on neuroticism may increase a learner's propensity to feel lonely and vice-versa (Abdellaoui et al., 2018). Loneliness can represent heightened intrinsic sensitivity to negative social stimuli, while neuroticism may reflect a heightened sensitivity to negative stimuli (Cacioppo et al., 2006).

Negative event

A significant and positive association was found between onset of negative events and neuroticism change ($t(100) = 2.362, p = .02$, partial $r = .23$). Due to the framing of the question, the results can be interpreted as *those individuals who perceived themselves to experience a greater number of negative events during their year abroad, became more neurotic over time*. This relationship is indicative of the notion that negative experiences will typically lead to negative emotions and if experienced continuously, this may lead the individual "to internalise this unpleasant affect" (Soto, 2015, p. 5), ultimately manifesting as an increase in neuroticism over time. It is also important to note

that sojourners negative reaction to stressors may be heightened because the support networks available at-home is no longer available. Moreover, cultural and linguistic differences can mean finding a resolution is made more difficult.

Participation in extracurricular activities

Participation in extra-curricular activities appeared to facilitate greater emotional stability over time, with a significant, and negative association found between the two constructs ($p = .008$, partial $r = -.27$). Within a school-based setting, participation in these activities have been linked to psychological adjustment, peer acceptance, problem coping and communication skills (e.g., Darling, Caldwell, & Smith, 2005; Fredrick & Eccles, 2006; Wilkinson & Hansen, 2006). This finding extends this relationship within a study abroad context, where adaptation and integration are required. Again, participation in these activities can be seen to facilitate the socially desirable trait-like changes and minimise the onset of stress and anxiety over time.

5.3.4.6 *Qualitative findings*

Participants who took part in the focus groups were also asked if there were any environmental triggers which could be associated with personal growth. Many who answered associated their personal growth with experiencing negative events as highlighted in the comments below.

“It is trying to take the positives out of the negative events and knowing you will develop from them. I know a lot of people who avoid going into difficult situations for fear of failure or for feeling disappointment. Ultimately, you will grow from these experiences and going abroad have made me appreciate this so much.” (participant 1)

“I think you grow the most from experiencing negative events. You change and develop, learning from those negative experiences. It sounds airy-fairy to say you grow so much as a person but what that actually means that there will be some really bad times which then mean you come out as a stronger person”. (participant 10)

“It is all about looking back, and you think if I managed to overcome the event and survive it, what else can I achieve and overcome. It is all your own threshold changing because it these events which now serve as your baseline.” (participant 2)

The above comments indicate the importance of encountering and overcoming adversity to stimulate personal growth positively. It is interesting to note, that all quantitative analysis, has tended to show negative events as being negatively associated with the particular outcome under

measure (e.g., broad trait change; well-being). This may indicate that while ‘*in the moment*’, negative events may lead to negative thoughts and emotions, retrospectively, individuals can reflect on and appreciate the occurrence of such events on their personal growth.

When asked if participants could give examples of such negative events, accommodation difficulties were often mentioned. However, by overcoming these difficulties, participants spoke of a sense of empowerment. Participant 10 highlighted how finding somewhere to live in adverse circumstance gave them the belief that they could overcome any other such adversity.

“I arrived in France, not having anywhere to live and booked an Airbnb. I hoped to find accommodation within a 6-day period, and I only found somewhere on the last day when I didn’t have anywhere to stay that night, so that was quite stressful. But then I knew that if I could do that, then I would be alright.” (participant 10)

Similarly, participant 9 mentioned how encountering accommodation difficulties made them take responsibility for their actions and acknowledge the need to overcome problems independently.

“My Airbnb was awful and dangerous. I had to leave, and I changed hostels four times, and then I ended up living in a hostel for 3 months because accommodation was so expensive and hard to get. So this was hard, but I accepted responsibility that I just couldn’t ring the university and expect them to sort it out.” (participant 9)

Consequently, an argument can be made that practitioners should better emphasise the facilitative capacity that encountering and overcoming negative events can have on personal growth. Whereas negative events may be currently attached with failure, sadness or anger, they should instead be seen as an opportunity for personal growth and should be embraced.

5.3.4.7 *Summary of findings*

Section 5.3.4 has looked to investigate why some sojourners experienced a greater change in their personality than others. By using change scores as the dependent variable, the regression models have been able to account for differences in the extent and directionality of trait-like tendencies over time. In brief, it can be said that loneliness inhibited perceived positive growth, while participation in extra-curricular activities fostered such growth. Furthermore, within the qualitative dataset, there was a consensus that upon reflection, overcoming negative events and experiences provided the strongest stimulant for positive growth abroad.

5.3.5 State personality and situational contingencies

Attention is now given to state personality; the most fine-grained level of the three levels of personality measured. Each state corresponded with a broad trait measured earlier in this study (i.e., openness, conscientiousness, extraversion, agreeableness and neuroticism). Capturing information on state personality is at the forefront of scholarly research, and given that perceived thoughts, feelings and behaviours are captured ‘*in the moment*’, the findings of this section hold the strongest ecological validity, as the scores have not been biased by retrospectivity.

Sample characteristics

Individuals who failed to complete more than 5% of all datapoints were dropped from all analysis below. Of the 151 individuals who remained, the response rate ranged from 5.95% to 92.86%. When exploring average scores, the fewer the data points, the less reliable the average score represents an individual’s typical personality. Multilevel models can, on the other hand, take this missing data into account (see section 5.2.7) and as such serves as a more appropriate approach when assessing intensive, repeated measures data.

5.3.5.1 *Descriptive statistics*

The descriptive statistics (Table 48) provided the mean, SD and range of mean scores for each measured state. The mean state score represents an average score across all 84 timepoints (i.e., an individual’s grand mean). The results showed *sojourners* to score on average higher in state openness ($d = .23$), conscientiousness ($d = .32$), extraversion ($d = .23$), and agreeableness ($d = .40$), while also scoring on average lower in neuroticism ($d = .53$) than their at-home peers on average across all timepoints. The SDs for each state were fairly large indicating individual differences in mean state scores, while the range column provided the lowest and highest average state score in each learning context (state personality means could range from 0 to 7).

Table 48: Descriptive State Personality Statistics (Means/SDs)

	Openness			Conscientiousness			Extraversion			Agreeableness			Neuroticism		
	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range
Sojourners (n = 92)	4.31	.76	2.63 - 6.76	4.72	.63	3.46 - 6.78	4.05	.56	3.16 - 6.37	5.16	.67	3.91 - 6.98	2.95	.72	1.09 - 4.88
Non- Sojourners (n = 59)	4.13	.74	2.73 - 6.43	4.50	.71	2.87 - 6.34	3.74	.68	2.18 - 4.83	4.89	.67	3.56 - 6.60	3.43	.81	1.72 - 5.25

5.3.5.2 The relationship between the broad traits and states

It was first ascertained whether baseline broad trait scores significantly predicted the average score of the corresponding state. Multilevel models were conducted, with state personality serving as the dependent variable and baseline trait score as the independent variable (Table 49).

Table 49: Parameter Estimates for Linear Growth Model of State Personality as a Function of Broad Traits

	State O		State C		State E		State A		State N	
	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2	Mod 1	Mod 2
Fixed effects										
Intercept	4.24*	4.24*	4.64*	4.64*	3.94*	3.94*	5.05*	5.06*	3.18*	3.18*
	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)
Trait O		0.43*		<.001		0.09		-0.02		-0.13
		(0.11)		(0.09)		(0.09)		(0.09)		(0.11)
Trait C		0.25*		0.53*		0.10		0.16		0.02
		(0.10)		(0.08)		(0.08)		(0.09)		(0.10)
Trait E		-0.16*		-0.03		0.12*		-0.06		<-.001
		(0.07)		(0.06)		(0.06)		(0.06)		(0.07)
Trait A		0.20*		0.20*		0.05		0.39*		-0.18
		(0.09)		(0.08)		(0.08)		(0.08)		(0.10)
Trait N		-0.05		-0.01		-0.19*		-0.15*		0.32*
		(0.07)		(0.06)		(0.06)		(0.06)		(0.07)
Random effects										
Intercept	0.54*	0.44*	0.41*	0.29*	0.35*	0.29*	0.44*	0.34*	0.55*	0.45*
	(0.73)	(0.66)	(0.82)	(0.53)	(0.59)	(0.54)	(0.66)	(0.58)	(0.74)	(0.67)
Residual	0.72*	0.71*	0.82*	0.82*	0.90*	0.89*	0.49*	0.49*	0.75*	0.75*
	(0.85)	(0.84)	(0.90)	(0.90)	(0.94)	(0.94)	(0.70)	(0.70)	(0.86)	(0.86)
ICC	.42		.33		.28		.47		.42	
Model fit statistics										
-2 Log Likelihood	11699	11595	12231	12117	12594	12479	10033	9929.8	11891	11801
<i>Note: O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism</i>										
Level 1 model: Stateopenness _{ij} = β _{0j} + β _{1j} (Trait openness) + e _{ij} , where Stateopenness _{ij} was the level of openness of person <i>j</i> on occasion <i>i</i> , β _{0j} refers to the average score of state openness across all occasions, β _{1j} was the regression coefficient of broad trait openness on state openness for person <i>j</i> , and e _{ij} was an error term. Level 2 model: β _{0i} = γ ₀₀ + u _{0i} and β _{1i} = γ ₁₀ + u _{1i} , where γ ₀₀ was the grand mean for openness across participants and occasions and γ ₁₀ was the mean of the standardized within-person regression coefficients of trait openness on state openness.										
* <i>p</i> < .05										

The pattern of results in Table 49 demonstrated that the corresponding trait served as the strongest predictor of state personality. For example, trait openness was the strongest predictor of state openness ($\beta = 0.43$, $SE = 0.11$, $p = <.05$). For example, the model estimates that being one-unit higher than the average on trait openness at baseline was associated with a state average openness score of 4.67. This finding has been replicated in other studies investigating state personality (Ching et al., 2014; Fleeson & Gallagher, 2009; Wilt, Nofhle, Fleeson, & Spain, 2012) and indicates that the state measure employed adequately captures the thoughts, feelings and behaviours associated with each trait.

5.3.5.3 Capturing variability across the personality levels

A unique aspect of this study is that personality has been explored at three levels (broad traits, narrow traits, and states). Analysis was guided by the following research question and examined data from both sojourners and non-sojourners:

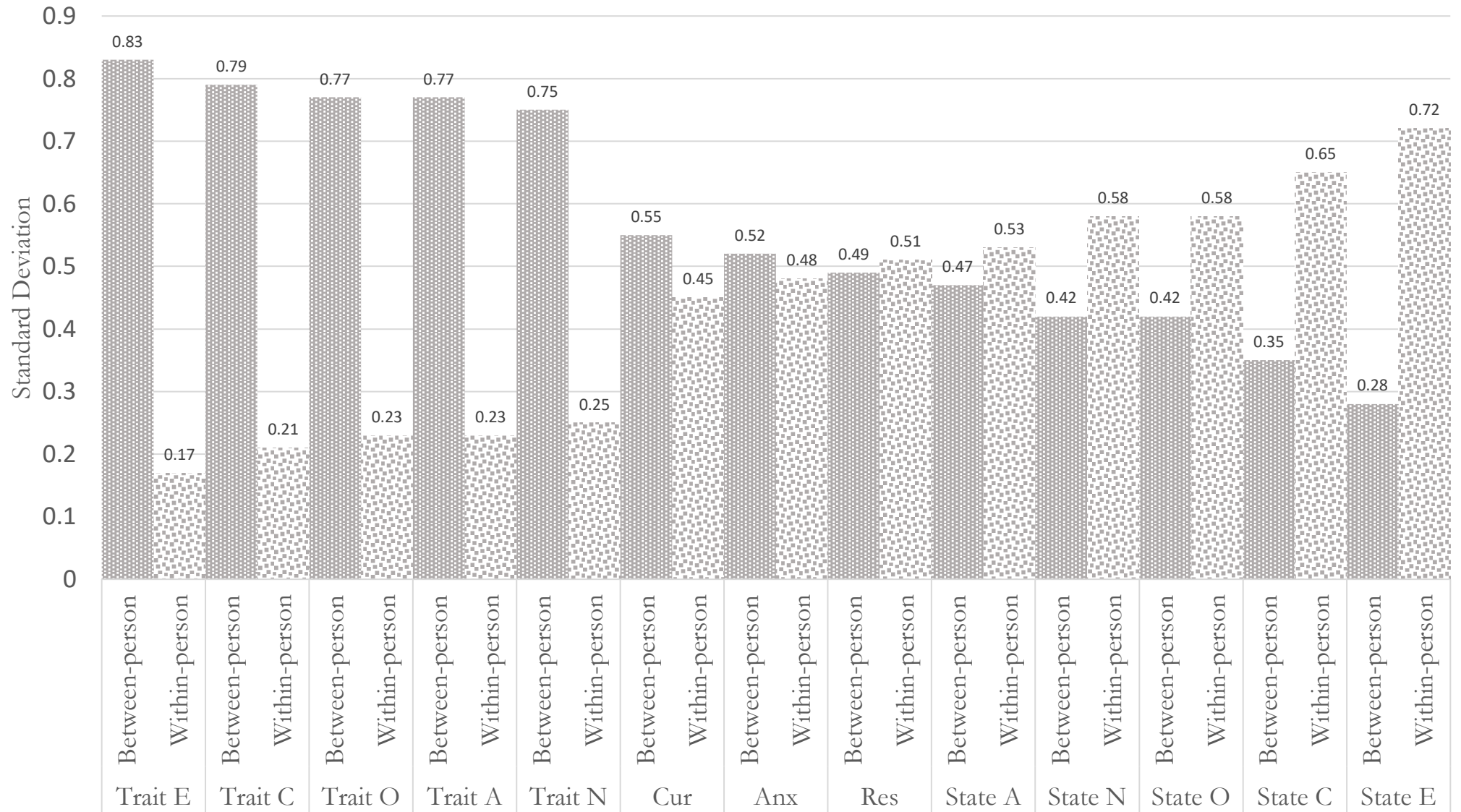
RQ5: What is the breakdown in variability at the three levels of personality

As shown in Table 50 and Figure 8, for the broad traits, much of the variability is found at the between-person level, ranging between 0.75 (neuroticism) and 0.83 (extraversion). These high ICC values indicate learners differ substantially between each other in their trait scores (i.e., individual differences) and that individuals show little fluctuation in their trait scores over time. This is to be expected, given that broad traits are relatively stable, and only show gradual change across the lifespan (e.g., Caspi, Roberts, & Shiner, 2005). For the narrow traits, variability at the within and between-person level appeared more balanced, with between-person variability ranging between 0.49 (resilience) and 0.55 (curiosity). For state personality, the majority of variability was seen at the within-person level. Here, between-person variability ranged between 0.28 (extraversion) and 0.47 (agreeableness), reflecting that individuals differed more in their own responses over a condensed period, than they did against one another. Put differently; state personality scores showed the highest levels of fluctuations around the one's own mean compared to the broad and narrow traits. This finding indicates two insights. Firstly, within variability is greater when the repeated measurement is more intense (frequency). Secondly, within-person variability is greater, when the instrumentation is concerned with day-to-day personality as opposed to general personality over an extended period (time window).

Table 50: Between and Within-person Variability at each Measured Personality Level

		Between	Within
Broad Traits	Openness	0.77	0.23
	Conscientiousness	0.79	0.21
	Extraversion	0.83	0.17
	Agreeableness	0.77	0.23
	Neuroticism	0.75	0.25
Narrow Traits	Anxiety	0.52	0.48
	Curiosity	0.55	0.45
	Resilience	0.49	0.51
State	Openness	0.42	0.58
	Conscientiousness	0.35	0.65
	Extraversion	0.28	0.72
	Agreeableness	0.47	0.53
	Neuroticism	0.42	0.58

Figure 8: Between and Within-Person Variability in Measured Personality Variables (across all learning contexts)



5.3.5.4 Systematicity of variability in state agreeableness

The following section looked to understand whether fluctuations in state personality can be described as error variance or whether such variability is dependent on the situation currently being experienced. All individuals were examined, with analyses guided by the question:

RQ6: How is variability in state agreeableness related to individuals' perception of situational characteristics?

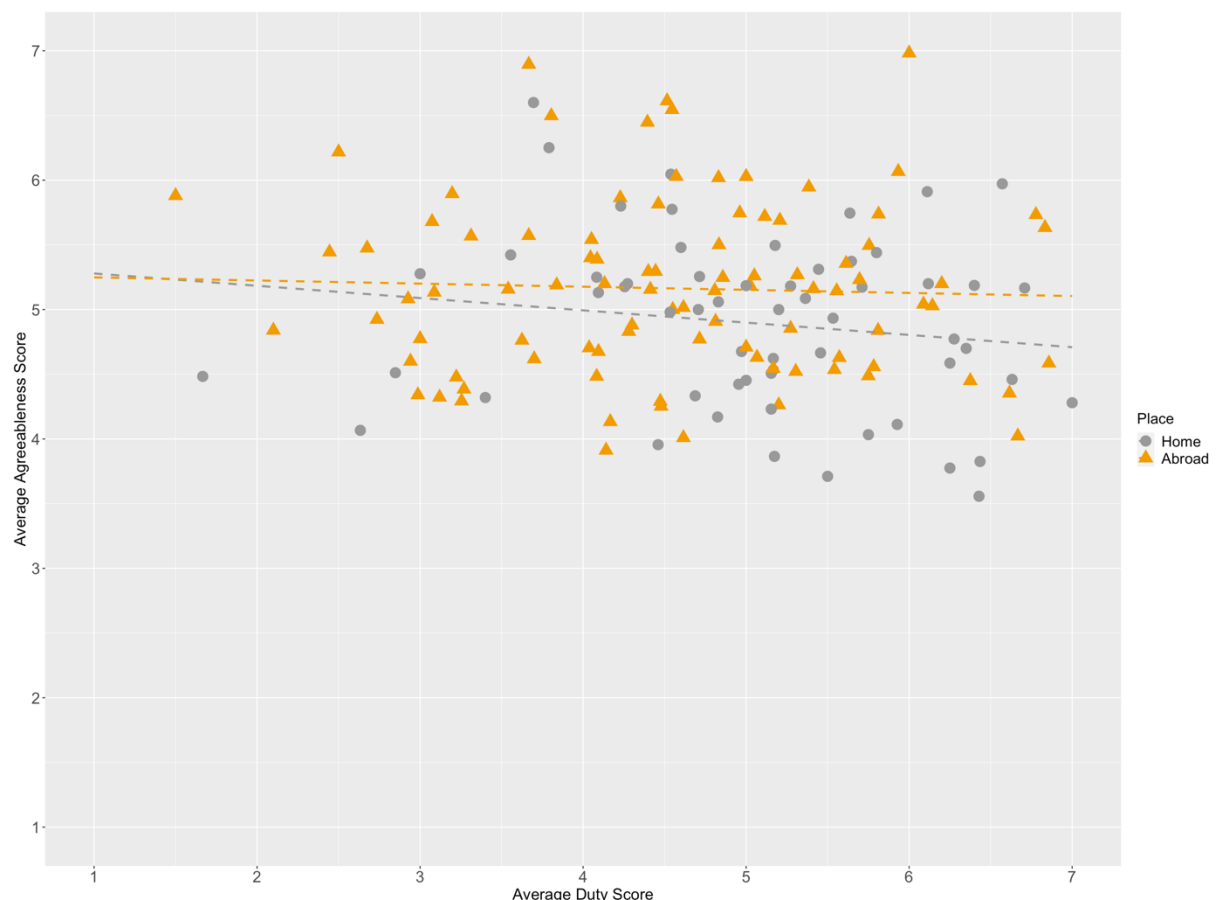
Here, attention was given to understanding the relationships between state agreeableness and a series of situational characteristics (e.g., duty, intellect, person familiarity and environmental security). The decision to choose state agreeableness as the dependent variable was empirically based, given that only trait agreeableness demonstrated a significant interaction in the earlier analysis (see section 5.3.2). The first of the multilevel models looked to ascertain whether learning context influenced the slope trajectories of agreeableness scores over time. To test between-group differences, a random slope, random intercept multilevel model was conducted.

Table 51: Parameter Estimates for Linear Growth Model of State Personality as a Function of Situation

	D	I	A	D	S	P.F	E.F	E.S	T.F	
Fixed effects										
Intercept	4.89*	4.89*	4.88*	4.88*	4.91*	4.98*	4.96*	4.89*	4.89*	(0.08)
Learning context	0.26*	0.26*	0.29*	0.29*	0.24*	0.24*	0.26*	0.26*	0.27	(0.11)
Situation Characteristic	-0.02*	-0.01	-0.22*	-0.17*	0.05*	.002*	-.001	.003*	<.001	(.001)
L.C * S.C	0.02*	0.02	-.001	-0.02	.001	<.001	<.001	.002	<-.001	(.001)
	(0.01)	(0.01)	(0.06)	(0.07)	(0.01)	(.001)	(.001)	(.001)		
Random effects										
Intercept	0.43*	0.42*	0.41*	0.40*	0.42*	0.38*	0.43*	0.40*	0.42*	(0.65)
Sit Char WP	.001*	.002*	0.05*	0.05*	.001*	<.001*	<.001*	<.001*	<.001*	(.004)
Residual	0.48*	0.48*	0.45*	0.47*	0.47*	0.48*	0.48*	0.48*	0.48*	(0.69)
	(0.69)	(0.69)	(0.67)	(0.69)	(0.68)	(0.69)	(0.69)	(0.69)		
Model fit statistics										
-2 LLH	9883.71	9927.65	9645.76	9838.51	9804.50	5790.00	9786.31	9852.63	9849.01	
<i>Note: D = Duty; I = Intellect; A = Adversity; D = Deception; S = Sociality; P.F = Person Familiarity; E.F = Environmental Familiarity; E.S = Environmental Security; T.F = Task Familiarity; L.C = learning context; S.C = situational characteristic</i>										
<i>Note: Level 1 model: StateAgreeableness_{it} = β_{0i} + β_{1i}(SituationCharacteristic)_{it} + e_{it}, where β_{1i} refers to the difference in average state agreeableness score between sojourners and non-sojourners across all time points for the i-th individual. Level 2 model: β_{0i} = γ₀₀ + γ₀₁^X(LearningContext)_i + u_{0i} where γ₀₀ is the average score of state agreeableness across time and γ₀₁ is how much higher or lower state agreeableness is over time in the treatment group AND β_{1i} = γ₁₀ + γ₁₁^X(LearningContext)_i + u_{1i} where γ₁₀ is the change in agreeableness per time point for the comparison group and γ₁₁ is the difference in the rate of change in the treatment group compared to the comparison group.</i>										
* p = <.05										

The findings (Table 51) indicated only duty to demonstrate a significant interaction effect ($\beta_4 = 0.02, p = .04$), with this relationship displayed in Figure 9, with each datapoint reflecting an individual. This interaction term suggests that a 1-unit increase in duty (i.e., a situation where work is required), sojourners perceived their state agreeableness to be 0.02 units higher on average than those at-home. That is to say; sojourners reported behaving significantly more agreeably when work was required compared to non-sojourners. A possible explanation to the finding is that *work* in the sojourning context required face-to-face contact or was teamwork orientated, whereas, at-home, work is likely to be academic, which is often completed alone. The literature has long pointed towards the symbiotic relationship between agreeableness, and teamwork-based activities (e.g., Driskell et al., 2006; Van Vianen & De Dreu, 2001) has found a positive relationship between agreeableness and teamwork. Regarding the remaining situations, no significant interaction effects were found: Intellect ($p = .16$), adversity ($p = .89$), deception ($p = .72$), sociality ($p = .96$), person familiarity ($p = .96$), environmental familiarity ($p = .83$), environmental security ($p = .19$) and task familiarity ($p = .63$).

Figure 9: Interaction Plot of Relationship Between State Agreeableness and State Duty across Learning Contexts



The systematicity of within-person variability was also explored (see Table 52) in each learning context separately. Here, multilevel models were conducted in order to ascertain the relationship between fluctuation in state personality and fluctuations in perceived situations characteristics, termed in the literature as ‘situation contingencies’ (Fleeson, 2001).

When exploring each learning context separately, state agreeableness was positively associated with the situational antecedents of sociality (Abroad: $\beta_1 = .05, p = <.001$; Home $\beta_1 = .05, p = <.001$), person familiarity (Abroad: $\beta_1 = .02, p = <.001$; Home $\beta_1 = .02, p = <.001$) and environmental security (Abroad: $\beta_1 = .05, p = <.001$; Home $\beta_1 = .03, p = <.001$). Conversely, adversity (Abroad: $\beta_1 = -.22, p = <.001$; Home $\beta_1 = -.22, p = <.001$) and deception (Abroad: $\beta_1 = -.20, p = <.001$; Home $\beta_1 = -.18, p = <.001$) were negatively associated with state agreeableness. Deception and adversity displayed the largest effect of within-person variability in agreeableness. Table 52 shows that state agreeableness was found to be significantly associated with duty in those at-home ($\beta_1 = -.02, p = .008$) but not in those abroad ($\beta_1 = .001, p = .94$).

There was also evidence of significant individual differences in the contingencies. To investigate these further, the range of slopes (β_1) into which fell 68% of the sample was estimated. The majority of both *sojourners and non-sojourners* adjusted their state agreeableness positively when perceiving positive changes in the variables of sociality, person familiarity and environmental security. Likewise, state agreeableness was negatively adjusted when responding to negative changes in adversity, deception, and environmental familiarity against one’s typical level of these variables. Nonetheless, these patterns were not always uniformed. For example, some sojourners became less agreeable when they were familiar with the people around them.

Table 52: Descriptive Statistics for Situation-Contingent Agreeableness Slopes Estimated by MLM

	Abroad (df = 92)				Home (df = 59)			
	γ_{10}	μ_1	Range of β_1	p-value	γ_{10}	μ_1	Range of β_1	p-value
Duty	.001	.04	$-.04 \leq \pi_1 \leq .04$.94	-.02	.01	$-.03 \leq \pi_1 \leq -.01$.008
Intellect	.004	.04	$-.39 \leq \pi_1 \leq .40$.65	-.01	.05	$-.06 \leq \pi_1 \leq .04$.12
Adversity	-.22	.24	$-.14 \leq \pi_1 \leq .02$	<.001	-.22	.22	$-.44 \leq \pi_1 \leq 0$	<.001
Deception	-.20	.28	$-.48 \leq \pi_1 \leq .08$	<.001	-.18	.15	$-.33 \leq \pi_1 \leq -.03$	<.001
Sociality	.05	.04	$.01 \leq \pi_1 \leq .09$	<.001	.05	.03	$.02 \leq \pi_1 \leq .08$	<.001
Person Familiarity	.02	.03	$-.01 \leq \pi_1 \leq .05$	<.001	.02	.02	$0 \leq \pi_1 \leq .04$	<.001
Env. Familiarity	-.01	.04	$-.05 \leq \pi_1 \leq .03$.09	-.01	.05	$-.06 \leq \pi_1 \leq .04$.18
Env. Security	.05	.07	$-.02 \leq \pi_1 \leq .13$	<.001	.03	.03	$0 \leq \pi_1 \leq .03$	<.001
Task Familiarity	-.01	.03	$-.04 \leq \pi_1 \leq .02$.75	.01	.05	$.04 \leq \pi_1 \leq .06$.70

Level 1 model: Stateagreeableness = $\beta_0 + \beta_1$ (e.g., Duty) + e, where β_1 refers to each individual's slope.

Level 2 model: $\beta_0 = \gamma_{00} + \mu_0$ and $\beta_1 = \gamma_{10} + \mu_1$, where γ_{10} refers to the grand mean of slope and μ_1 refers to the deviation of individuals' slopes from the mean slope.

Note: All variables are contingent with state agreeableness

Table 52 has provided evidence that variability in within-person state agreeableness is systematic to changes in the perceived situation characteristics and therefore, lends support to the notion of situational contingencies.

5.3.5.5 Summary of findings

Section 5.3.5 has examined a) the existence and differences in between and within-person variability at the three observed levels of personality (broad, narrow, & state) and b) whether variability in state agreeableness is related to the selected situational cues and characteristics.

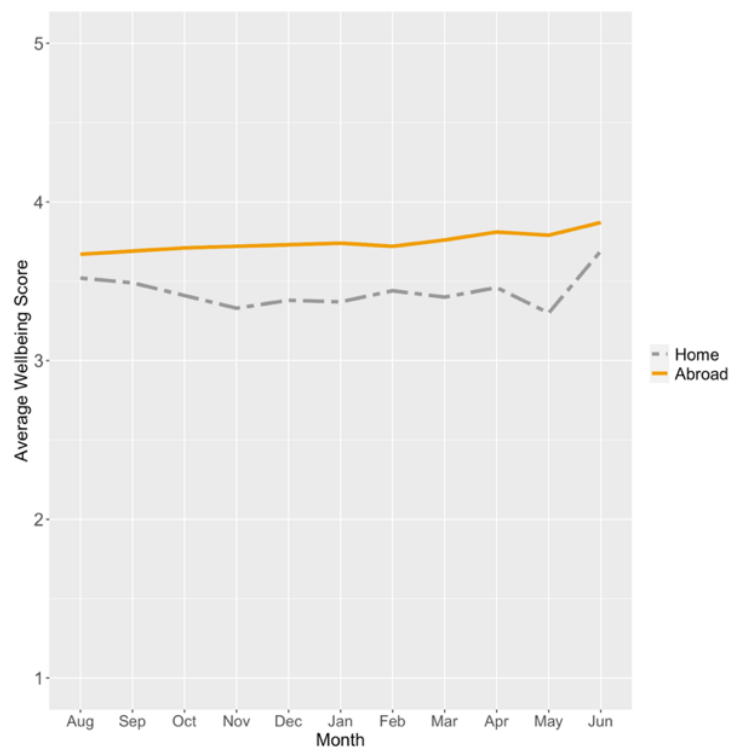
The results have indicated that the more intensive the repeated measure is; the greater amount that total variability is accounted for by within-person variability. That is to say, at the broad trait level, individuals tend to be fairly similar in their own scores over time and differ greatly between each other. Conversely, at the state level, the same individual differs greatly around their own mean, and consequently, the between-person differences are not as great. This section has also found that within-person variability is systematically associated with the relationship between personality and the situation (e.g., duty, sociality). Of particular interest to this analysis is the finding that Duty was significantly related to the learning context. Here, it was found that sojourners reported behaving significantly more agreeably in a situation which was perceived to require more work than non-sojourners. Together, these findings support the notion that within-personality variability should not be viewed as error variance and warrants further investigation.

5.3.6 Capturing well-being change and its predictors

As shown in Figure 10, *sojourners* tended to show an incline in well-being over the year, while *non-sojourners* tended to experience declining well-being over the year, rebounding post-exams. Baseline scores indicated that well-being was not significantly higher in one learning context over another as evidenced by an independent samples t-test, $t(154) = -1.562, p = .120$. In order to ascertain whether fluctuations in well-being were meaningful, multilevel model analysis was undertaken. The model estimated that for *sojourners*, well-being was significantly higher in the months of April ($\beta_8 = 0.12, p = 0.03$), May ($\beta_9 = 0.12, p = 0.04$) and June ($\beta_{10} = 0.17, p = .003$) compared to baseline score of $\beta_0 = 3.67$, achieved in August. Therefore, it could be said that well-being was highest when, for many, the experience was coming to an end. The gradual incline does, however, suggest a cumulation effect over time, rather than perhaps general relief that the experience is nearly over. It found for those at-home, when compared to an estimated average August score of $\beta_0 = 3.54$, well-being was significantly lower in the months of November ($\beta_3 = -0.21, p = .002$), December ($\beta_4 = -0.15, p = 0.03$), January ($\beta_5 = -0.17, p = .02$) and lastly May ($\beta_6 = -0.26, p = <.001$). These higher scores possibly coincide with when academic pressure is expected to be at its highest (e.g., dissertation/revision/exams).

Figure 10: Overview of Average Well-being Scores across Time and Learning Context

Time	Abroad			Home		
	N	Mean	SD	N	Mean	SD
Aug	97	3.67	.53	59	3.52	.66
Sept	104	3.69	.60	68	3.49	.66
Oct	103	3.71	.57	67	3.41	.57
Nov	107	3.72	.63	69	3.33	.67
Dec	109	3.73	.66	69	3.38	.69
Jan	104	3.74	.58	67	3.37	.66
Feb	108	3.72	.61	68	3.44	.56
Mar	108	3.76	.57	63	3.40	.64
Apr	105	3.81	.57	56	3.46	.68
May	105	3.79	.61	61	3.30	.70
Jun	100	3.87	.54	61	3.69	.60



5.3.6.1 Understanding the role of context on psychological well-being

The first of the analyses concerned ascertaining whether learning context had a direct effect on average well-being over time. Analysis was guided by the following research question and examined data from both sojourners and non-sojourners:

RQ7: Do sojourners experience significantly greater well-being over time compared to non-sojourners?

A comparison of well-being trajectories in each learning context was calculated by undertaking a random intercept, random slope multilevel model (Table 53). In this model, an interaction term between learning context and time was tested as per previous multilevel models (section 5.3.2).

Table 53: Parameter Estimates for Linear Growth Model of Well-being as a Function of Learning Context

	Model A (Null Model)	Model B (+ Time)	Model C (+ Learning Context)	Model D (+ Time * Learning Context)
Fixed effects				
Intercept	3.62* (0.03)	3.58* (0.04)	3.39* (0.06)	3.42* (0.06)
Time		0.01* (0.004)	0.01* (0.003)	.004 (.006)
Learning Context			0.30* (0.07)	0.25* (0.08)
Time * Learning context				0.01 (.008)
Random effects				
Intercept	0.23* (0.47)	0.23* (0.48)	0.22* (0.47)	0.22* (0.47)
Time		.0006* (0.02)	.0006* (0.02)	.0006* (0.02)
Residual	0.17* (0.41)	0.17* (0.42)	0.17* (0.41)	0.17* (0.41)
ICC	.58			
Model fit statistics				
-2 Log Likelihood	2438.62	2292.88	2276.14	2274.67

Note: Level 1 model: $\text{well-being}_{it} = \beta_{0i} + \beta_{1i}(\text{TIME})_{it} + e_{it}$, where β_{1i} refers to the average change in well-being score for the i -th individual over time. Level 2 model: $\beta_{0i} = \gamma_{00} + \gamma_{01} \times (\text{LEARNING CONTEXT})_i + u_{0i}$ where γ_{00} is the average score of well-being at the initial timepoint and γ_{01} is how much higher or lower the well-being score is at baseline in the *sojourn group* AND $\beta_{1i} = \gamma_{10} + \gamma_{11} \times (\text{LEARNING CONTEXT})_i + u_{1i}$ where γ_{10} is the change in well-being per timepoint for the *non-sojourn group* and γ_{11} is how much higher or lower the rate of change is in the *sojourn group* compared to the *non-sojourn group*. * $p = <.05$

The null model (Model A) indicated that the average well-being score across all timepoints and individuals was $\beta_0 = 3.62$. As ICC value of .58 (58%) indicated, the majority of variance was accounted for by differences in individuals (i.e., individual differences), while the remaining variability was accounted for at the within-person level. Model D serves as the full model and can be interpreted as follows. The average baseline score for the non-sojourn group was $\beta_0 = 3.42$ and was expected to demonstrate an increase in well-being over time by $\beta_1 = .004$ units. The model estimated sojourners to score $\beta_3 = 0.25$ units higher in well-being at baseline. The

interaction term between time and learning context was found to be non-significant ($\beta_4 = 0.01, p = .19$)⁸. As such, when exploring well-being scores between August and June, sojourning was not found to result in significantly higher well-being than remaining at-home.

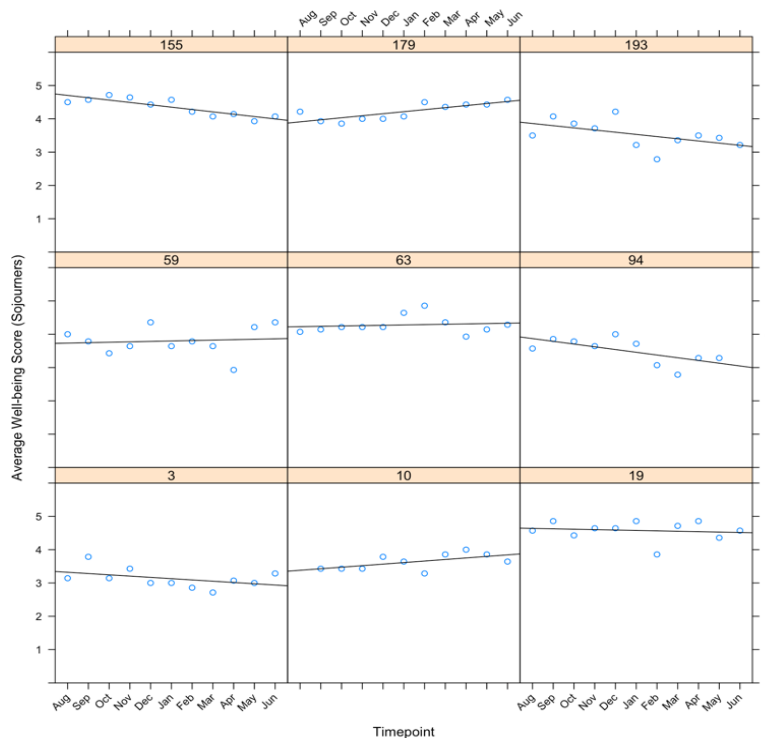
5.3.6.2 Well-being change over time

This section aims to ascertain a) whether sojourners experienced significant well-being change over time and b) whether sojourners differed in their well-being trajectories over time, with analysis guided by the following research question, examining the sojourn group only.

RQ8: Do sojourners experience significantly higher well-being over time, and is this uniform across all sojourners?

As shown in Figure 11, some individuals started with higher well-being than others, and some demonstrated an upward trajectory in scores, while others showed a decline. Moreover, some individuals displayed greater variability in scores over time than others.

Figure 11: Panel Plots of Well-being Scores across Time



*each number refers to a participant who was randomly chosen for the purpose of this figure

⁸ When taking into account only August to May, the interaction term became significant, with *sojourners* estimated to score $\beta_4 = .02$ units ($p = <.05$) higher in well-being at each timepoint compared to at-home. This significant term may not be a result of the sojourn experience itself, but rather that 3rd year undergraduate students may experience increasing pressure (e.g., academic, professional) throughout the academic year which may negatively impact well-being.

Given these differing trajectories, the most sensible approach was to conduct a random-intercept, random slope multilevel model. Such an approach sees each individual as having a different intercept (i.e., starting point) and a different slope (scores over time). This approach to multilevel modelling can account for these differences, with the output presented in Table 54.

Table 54: Parameter Estimates for Linear Growth Model of Well-being as a Function of Time

	Model 1A	Model 1B
Fixed effects		
Intercept	3.75* (0.04)	3.68* (0.05)
Time		0.01 (.005)
Random effects		
Intercept	0.19* (0.40)	0.26* (0.51)
Time		.002* (0.04)
Residual	0.16* (0.42)	0.13* (0.36)
ICC	.54	
Model fit statistics		
-2 Log Likelihood	1463.92	1377.31
<i>Note:</i> Level 1 model: $\text{well-being}_{it} = \beta_{0i} + \beta_{1i}(\text{TIME})_{it} + e_{it}$, where β_{1i} refers to the average change in well-being for the i -th individual over time. Level 2 model: $\beta_{0i} = \gamma_{00} + u_{0i}$ and $\beta_{1i} = \gamma_{10} + u_{1i}$, where γ_{00} is the average score of well-being at the initial timepoint and γ_{10} is the average monthly change in well-being.		
* $p < .05$		

Model 1A served as the *null model* and estimated the average well-being scores in sojourners across all timepoints to be $\beta_0 = 3.75$. The ICC value of .54 indicated that 54% of the variability in well-being scores was at the between-person level. Model 2B served as the full model, with the average baseline well-being score being estimated at $\beta_0 = 3.68$. Although Figure 10 displayed an upward trajectory over time, the model estimated sojourners to demonstrate a non-significant change in well-being over time ($\beta_1 = 0.01, p = .07$). Concerning the random effects, the intercept (0.26) and slope (0.02) values were significant and the magnitude of these calculated using the range of intercepts and slopes which fell in the middle 95% of learners. The model estimated that 95% of sojourners begun the year abroad with a well-being score between 2.70 and 4.66, and the extent of change at each timepoint ranged between -.07 and .08. Put differently, while some individuals experienced positive changes in well-being over time, others experienced a substantial decrease, reflecting those trajectories found in Figure 11. Lastly, the estimated residual was 0.13 and indicated that at any given moment, sojourners deviated from their own trajectory by up to approximately 0.70 units at any given timepoint and considered significant ($p = <.05$).

5.3.6.3 Predictors of fluctuations in psychological well-being

This section aims to analyse fluctuations in well-being in relation to fluctuations in the observed narrow traits. Analysis captures variability at both the between-and-within person level, and was guided by the following research question, examining sojourner data only.

RQ9: Are monthly well-being scores contingent on monthly narrow personality scores?

Earlier analysis (5.3.6) showed sojourners well-being to fluctuate over time. That is to say, at any given point in time, an individual's well-being may be significantly higher or lower than the previously measured point. The purpose of the following analysis is to identify the relationship between these fluctuations and fluctuations in the observed narrow traits (i.e., whether well-being scores were contingent on narrow trait scores). In doing so, it becomes possible to better understand the variables which promote positive well-being. Previous literature has tended to view well-being as holistic and dependent on more objective cues in the immediate environment (e.g., social networking). Conversely, little thought has been given to understanding the mediating role the narrow traits (e.g., curiosity, anxiety, resilience) may have on strengthening or weakening positive well-being in a sojourning environment (Alharbi & Smith, 2019; Russell, Thomson, & Rosenthal, 2008). To explore this, a series of multilevel models were conducted.

Table 55: Monthly Well-being Scores as a Function of Monthly Narrow Trait Scores (sojourners)

	Anxiety		Curiosity		Resilience	
	Model 1A	Model 1B	Model 1A	Model 1B	Model 1A	Model 1B
Fixed effects						
Intercept	3.75* (0.04)	3.75* (0.02)	3.75* (0.04)	3.75* (0.03)	3.75* (0.04)	3.75* (0.02)
Trait Between		-0.73* (0.03)		0.43* (0.02)		0.62* (0.04)
Trait Within		-0.53* (0.01)		0.35* (0.02)		0.43* (0.02)
Random effects						
Intercept	0.19* (0.40)	0.03* (0.19)	0.19* (0.40)	0.13* (0.36)	0.19* (0.40)	0.07* (0.27)
Trait Within		0.01* (0.10)		0.03* (0.18)		0.01* (0.13)
Residual	0.16* (0.42)	0.06* (0.25)	0.16* (0.42)	0.11* (0.33)	0.16* (0.42)	0.08* (0.29)
ICC	.54		.54		.54	
Model fit statistics						
-2 Log Likelihood	1463.92	410.16	1463.92	1086.19	1463.92	775.45
<p>Note: Level 1 model: $\text{well-being}_{it} = \beta_{0i} + \beta_{1i}(\text{resilience}_{it}) + e_{it}$, where β_{1i} refers to the average resilience score for the i-th individual when the i-th individual is at their typical level of resilience. Resilience has been person-mean centered so that 0 on this variable is the typical resilience score for each individual. Level 2 model: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{resilience}_i) + \mu_{0i}$ where γ_{00} is the grand mean of resilience for the sample, γ_{01} is the effect of the average resilience score for the i-th individual on the sample mean, and μ_{0i} is the well-being mean for each individual AND $\beta_{1i} = \gamma_{10} + \mu_{1i}$ where γ_{10} is the average within-person relationship, and μ_{1i} is the association between resilience and well-being for the i-th individual.</p>						
* $p < .05$						

5.3.6.3.1 *Anxiety*

Table 55 first presents the multilevel model for anxiety. For *sojourners*, Model 1B indicated that the mean anxiety score across all months was 3.75 (95% of students fall within a range of 3.41 and 4.08). At the between-person level, people who had a higher average anxiety score had a lower average well-being score. A person's average well-being was predicted to be 0.73 units lower for each 1-unit increase in average anxiety. At the within-person level, when anxiety was 1-unit above their own typical anxiety score, their well-being score was predicted to be 0.53 units lower than their own typical well-being. The range of slopes for 95% of students fell between -.33 and -.72, indicating that for the majority of sojourners, their well-being was lower when anxiety was higher, although the strength of this relationship differed among individuals.

5.3.6.3.2 *Curiosity*

The output of Table 55 indicated that at the between-person level, a *sojourner's* well-being score was estimated to be, on average, 0.43 units higher for each 1 unit increase in curiosity. Consequently, those who perceived themselves to be higher in curiosity over time tended to have higher well-being than less curious individuals. At the within-person level, a 1 unit increase in a participant's own mean curiosity level, lead to a 0.35 unit increase in their own well-being. 95% of sojourners fell within a range of 0 and 0.68, indicating that for a minority of students, an increase in average curiosity did not necessarily lead to higher well-being for those abroad.

5.3.6.3.3 *Resilience*

Lastly, resilience was explored and demonstrated strong associations with well-being scores in each learning context (see Table 55). At the between-person level, *sojourners* who have a higher mean resilience also have a higher mean well-being. A person's average well-being was estimated to be 0.62 units higher for each 1 unit increase in their average resilience. At the within-person level, when resilience was 1 unit above an individual's own mean resilience score, their well-being was estimated to be 0.43 units higher, with the range of slopes for 95% of the sojourners falling between 0.41 and 0.63.

In sum, fluctuations in well-being were shown to be contingent on fluctuations in the three measured narrow traits and therefore indicates there is value in looking to take a bottom-up approach to improve well-being from a stakeholder perspective as opposed to looking to manipulate 'well-being' as a holistic entity by taking a top-down approach.

5.3.6.4 Predictors of average psychological well-being across the year

While section 5.3.6.3 has explored fluctuations in well-being during the year abroad, this section looks to examine the association between overall average well-being and the *person* and *environmental* variables collected. Consequently, the following analysis looks only at the between-person level, with analysis conducted on the sojourner sample guided by the following question:

RQ10: What is the relationship between average well-being scores and the captured variables?

Conclusions drawn are supported by both quantitative and qualitative data. Firstly, regression models within the domain of the *person* and the *environment* were conducted. Given the available sample size ($n = 110$), no more than six predictors could be added to a model at one time. Two sojourners failed to complete the broad trait instrument at baseline and were as such dropped from the regression output found in Table 56. All cases ($n = 110$) were used when assessing factors associated with the environment (Table 57).

5.3.6.4.1 Factors associated with the person

Table 56 below provides the output of the multiple regression analysis in which an individuals' average well-being score across the year was regressed on baseline broad personality trait scores. Concerning assumptions, partial regression plots indicated linearity. A Durbin-Watson statistic of 2.153 indicated independence of residuals. There was homoscedasticity and no evidence of multicollinearity. There were no outliers, and the assumption of normality was met.

Table 56: Multiple Regression of Average Well-being Scores on Baseline Broad Traits

Factors	Unstan. Coefficients		Standardised coefficients			zero-order	partial
	B	SE	Beta	T	Sig.		
Constant	3.198 (2.281, 4.115)	.462		6.917	<.001		
Openness T1	.126 (-.019, .271)	.073	.148	1.730	.087	.134	.169
Conscientiousness T1	.140 (.007, .274)	.067	.186	2.082	.040	.305	.202
Extraversion T1	.023 (-.071, .118)	.048	.045	.491	.625	.232	.049
Agreeableness T1	.048 (-.095, .191)	.072	.060	.667	.506	.194	.066
Neuroticism T1	-.225 (-.327, -.122)	.052	-.395	-4.345	<.001	-.449	-.395

$R^2 = .276$, Adjusted $R^2 = .240$, $F(5, 102) = 7.762$, $p < .001$

The results of Table 56 indicated that baseline conscientiousness was significantly and positively related with average well-being ($p = .04$, partial $r = .20$), while neuroticism was significantly, and negatively associated with higher well-being ($p < .001$, partial $r = -.39$). Baseline broad personality traits were shown to explain 24% of the variability in average well-being scores.

5.3.6.4.2 *Factors associated with the environment*

Table 57 presents the regression output for variables associated with the environment. Regarding assumptions, linearity was assessed by partial regression plots and a plot of studentised residuals against the predicted values. Independence of residuals was achieved as assessed by a Durbin-Watson statistic of 2.083. There was homoscedasticity, and no evidence of multicollinearity as all variables had a tolerance value of greater than 0.1. There were no outliers, and the assumption of normality was met, as assessed by a Q-Q Plot.

Table 57: Multiple Regression of Average Well-being Scores on Environmental Variables

Factors	Unstan. Coefficients		Standardised coefficients				
	B	SE	Beta	T	Sig.	zero-order	partial
Constant	3.600 (3.190, 4.010)	.207		17.416	<.001		
Loneliness	-.543 (-.832, -.255)	.145	-.314	-3.738	<.001	-.455	-.346
Close Bond	.053 (-.367, .261)	.158	.029	.336	.738	.062	.033
Belongingness	.610 (.229, .991)	.192	.267	3.177	.002	.349	.299
Negative Event	-.458 (-.839, -.078)	.192	-.202	-2.390	.019	-.359	-.229
Participation	.285 (.063, .508)	.112	.203	2.540	.013	.279	.243
Travelled	.254 (-.052, .560)	.154	.128	1.646	.103	.150	.160

$R^2 = .383$, Adjusted $R^2 = .347$, $F(6, 103) = 10.655$, $p = <.001$

The output (Table 57) demonstrated that loneliness (partial $r = -.34$) served as the strongest predictor of poorer well-being abroad, while the onset of negative events was also associated with lower well-being (partial $r = -.22$). Conversely, belongingness (partial $r = .29$) and participation in extra-curricular activities (partial $r = .24$) were both significant predictors of higher well-being when holding all other variables constant. Creating a new friendship (partial $r = .03$) and travelling (partial $r = .16$) did not appear to facilitate higher well-being over the academic year. Given that the analysis has focused on an average score, it could also be argued that those who did perceive higher well-being also felt less lonely, perceived fewer negative events, felt part of the community and partook in extracurricular activities. Put differently, this analysis has explored association only, and no causal directionality can be given. For example, it cannot be determined whether feeling less lonely resulted in higher well-being, or whether high well-being resulted in lessened perceived loneliness. Nonetheless, variables associated with the environment were shown to explain 34% of the variability in average well-being scores.

5.3.6.5 *Qualitative findings*

To supplement the findings above, focus group participants were asked to describe how they perceived their well-being during the year abroad and identify drivers or barriers of well-being.

Academic pressure

Several individuals noted how they perceived their well-being to be higher during the year abroad, as sojourners felt less academic pressure than at-home. This decrease in academic rigour may be seen as encouragement from home universities for sojourners to explore culture and gain new experiences. After all, in the UK, many universities set a pass rate of 40% for their sojourners, with these marks not being carried forward to the final degree classification. It should be noted that these comments are only generalisable to those who studied.

It was lovely just to get out of the university treadmill. My friends were showing themselves working in the library and handing in essays, and I felt like I was on a mini holiday. (participant 7)

“I think strangely for me because I was outside my home university; there wasn’t that aspect of stress. In Russia, I was at university, but we barely had any kind of homework stuff, just some Russian language stuff and the two essays I left until the end. So actually, I would say that my well-being was better abroad than at my home university because I didn’t have that constant feeling that I should be doing work.” (participant 5)

“I also found it less stressful abroad. Exams were split in January and July; it felt less work than if we had to wait to do all exams in the end, like at my university.” (participant 6)

Boredom

Several participants chose to highlight how, as time passed, they found themselves becoming increasingly bored as the general excitement to try new things and meet new people dissipated. Boredom has typically been associated with negative thoughts, feelings, and behaviours, such as anxiety, loneliness and hopelessness (Elpidorou, 2017). The views of these individuals may be reflective of the extended stay (year-long), in the same location, for very little literature has discussed the onset of boredom in long-term sojourn stays. Taguchi (2015) found sojourners to remain stimulated during a short stay, but when asked if sojourners would have liked to stay for longer, there were fears of becoming bored. Nonetheless, Allen (2010) found that individuals

nearing the end of a six-week experience abroad to become increasingly bored, indicating boredom may not be specifically related to the length of stay.

“I became more restless near the end of my year abroad and found myself increasing in anxiety.” (participant 11)

“I didn’t really meet any new people in the second semester, and everyone’s excitement to travel had lessened. There was less happening, less excitement to see new places and a lot more people from home came to visit which meant I was doing the same sort of trips every weekend with different people which was quite draining.” (participant 5)

“I think at the start I was very busy, and I was getting used to everything, and I had a lot of stuff to do. I would say however that as time was passing, I had done all the things I wanted to, and I wasn’t meeting any new people anymore. I feel that I was getting bored and with it depressed” (participant 3)

Mid-term dip

In each of the focus groups conducted, participants noted how they found the second semester more psychologically challenging than the first, and in particular the return to their host country after Christmas. Participant 8 highlighted, in particular the contradictory state of their well-being across the year, with well-being being perceived to be considerably higher in the first three months. This was a sentiment shared by other sojourners, who noted that negative feelings such as loneliness and depression and anxiety, were more prevalent in the second semester caused predominantly by the loss of friendship groups and the lack of motivation to create new ones. It would appear, therefore, that the period after Christmas required re-adaptation (Beaven & Spencer-Oatey, 2016). This finding is not generalisable to programmes of one semester in length, which many exchange programmes are. These comments also suggest that the higher well-being scores found for April, May and June (see section 5.3.6) may have indeed been a result of a general excitement or relief to return home.

“For me, because I was there for the whole year, most of my friends who I was close with during the 1st semester, including my housemates, all left and as such there was a really rough period between January and March where there was a really lonely transition between finding new friends but being in the same place. It wasn’t the fact that I had changed country and therefore trying to find new friends. It was just as if they had all left me, which was quite hard to process and learning how to cope with loneliness during this time.” (participant 8)

I think for me, the second half of my year because all my friends left, I became more isolated in my room and I stayed in my room and didn't go out as much and do all the things I had expected to do. I think this triggered more anxiety because you are not doing what you think you should be doing, for example, integrating more."

(participant 13)

"I found the first three months, apart from the very initial anxiety, went really fast, and I found the first three months as really positive for my well-being. However, the second semester was so much harder. All the friends I had met in the first semester were now leaving, and this was more of a low point. There was a fallout from Christmas, it was cold, and most people were either studying for exams or were leaving." (participant 3)

Matching expectations

The last of the themes to be analysed was the discrepancy between expected experiences and realised experiences inhibited overall well-being. Participant 1 chose to highlight how they found it difficult to comprehend why the experience was so different from the one they had imagined. This was made more difficult by the realisation that their support networks were all having such different experiences. Similarly, participant 5 noted how it can be easy to blame oneself when the experience does not match what you were expecting. The available literature has consistently noted how having unrealistic expectations of study abroad can lead to negative outcomes including disappointment, depression, and ultimately disillusionment with culture and language (Engle & Engle, 2003; McLeod & Wainwright, 2009; Mendelsohn, 2004; Wilkinson, 1998).

"I struggled with the contrast between people seeing a year abroad as a long holiday, and you are going to have a great time and then feeling like I wasn't and subsequently the isolation caused by that because I had people at-home telling me it would be ok, I had a friend in Italy having a great time and who was telling me how wonderful it is and constantly telling me I should be having a great time which made it worse. I just felt like I should have been having a great time constantly, but I wasn't, and this made things more difficult to handle." (participant 1)

"I feel like the year abroad promo can be a bit like an Instagram highlight reel, and for me, it wasn't like that. Sometimes, when you aren't having the best time on your year abroad you are a bit like ... oh this is wrong; this doesn't match up what I have read about." (participant 5)

5.3.6.6 *Summary of findings*

Section 5.3.6 has evaluated whether a) sojourning has a facilitative effect on well-being and b) identified predictors of well-being at both the within and between-person level on a year abroad.

The findings have indicated that the average sojourner does not experience significantly higher well-being than those who remain at-home. Similarly, the average sojourner was not found to experience significantly higher well-being when abroad. When looking at predictors, there was evidence that well-being fluctuated in relation to fluctuations in the observed narrow traits. Moreover, those individuals who perceived themselves to be less lonely, experience fewer negative events and felt a member of the host community had on average higher well-being. During the focus groups, individuals spoke of becoming bored, and a mismatch in expectations served as antecedents of negative well-being abroad.

5.3.7 Language change and its predictors

5.3.7.1 Observed Language change in learners across the year

Section 5.3.7.1 examines proficiency changes in learners of French, German, Italian and Spanish during the year abroad, as operationalised by a C-test, with analysis guided by the following research question, examining data collected from language learners from the sojourn sample:

RQ11: Do learners return home with higher linguistic proficiency after a year abroad?

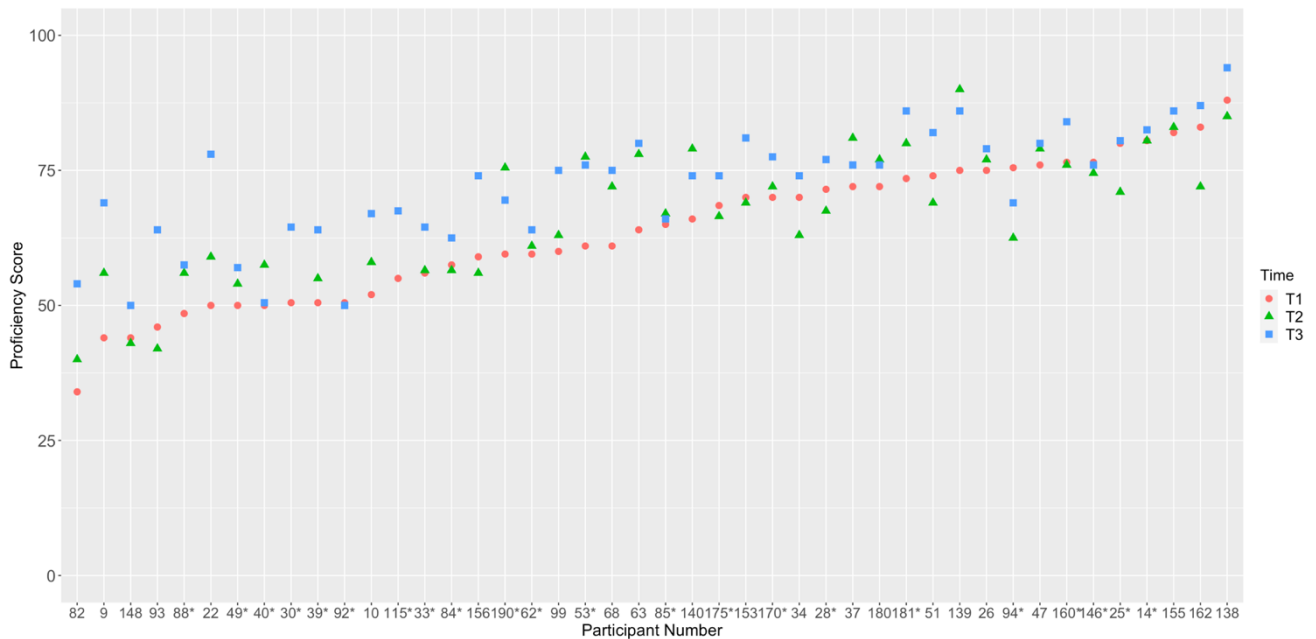
Due to the lack of a stay at-home control group, the findings can be described as exploratory only and do not infer a direct effect of learning context on facilitating language development. This is because it is unknown if a similar change trajectory would have been found in the comparison group. Analysis also considered whether learners perceived change in the four skills of listening, reading, speaking and writing. This analysis was supported by insights gathered from qualitative data, which examined in more detail the perceived changes of language learners.

All analyses below were conducted on a sample of 44 language learners, of whom were divided by Honours Programme. In the first group ($n = 22$) were *Joint Honour* (JH) students, who learnt two L2s and split their year abroad between two different host countries. In the second group ($n = 22$) were *Single Honour* (SH) students, who learn only one L2, with the majority remaining in the same host country for the duration abroad.

At each of the three timepoints (pre-sojourn, mid-sojourn, post-sojourn), a proficiency score was calculated. While it would have been preferable to analyse each language separately, given the small sample size, such analyses would have been underpowered. As such, a language score was created for each learner at each timepoint which was representative of learner proficiency, regardless of language learnt. For *Joint Honour* students, a proficiency score was calculated by averaging the language score achieved in each of their learnt languages. At baseline, a paired samples t-test found a non-significant difference in average proficiency between the first L2 ($m = 61.95$, $sd = 14.79$) and the second L2 ($m = 64.04$, $sd = 11.48$): $t(21) = .673$, $p = .508$, indicating these learners were not more proficient in one language over another. For *Single Honour* learners, this averaging was not necessary, and their language score at each timepoint was representative of their own ability in the specific language. A one-way ANOVA was conducted at baseline in order to ascertain whether learners of one language were significantly more proficient than that of another. This result was non-significant ($F(3, 18) = 2.061$, $p = .14$), indicating the language score found at baseline was representative of a learner's ability in any of the four languages.

Figure 12 displays the individual proficiency scores of the 44 individual participants at each timepoint, ordered by baseline proficiency scores. Each shape is representative of their linguistic ability, regardless of the language learnt. It shows that while almost all learners made gains in their proficiency over time, the higher achievers at baseline tended to make less improvement than those who scored in middle and low ranges at baseline. A paired samples t-test indicated learners significantly improved in their learnt language(s) by an average of 8.59 marks ($sd = 6.90$, min.: -6.50, max.: 28) from baseline to post-test: $t(43) = 8.251, p = <.001, d = .80$, which according to Cohen (1988) is a large-sized effect. Consequently, it can be said that language learners returned home significantly more proficient than when they departed for the year abroad. However, this statistic belies the differences in sample make-up, half of whom were *Joint Honour* and half of whom were *Single Honour* learners.

Figure 12: Average Proficiency Scores for Language Learners across Time



*Individuals with an asterisk are Joint Honour language learners

Note: Each shape is representative of linguistic proficiency in any of the four tested languages (French, German, Italian & Spanish).

Note: Participants are ordered by proficiency score achieved at pre-sojourn (baseline)

A two-way mixed model ANOVA was run in order to determine whether *Honours programme* (i.e., Single vs Joint Honours) had a significant effect on language change over *time*. The assumptions for this test were fully met. The findings indicated a statistically significant interaction between Honours programme and time on proficiency scores: $F(2, 78) = 3.443, p = .03$, partial $\eta^2 = .08$ (Figure 13). As the interaction was significant, further simple main effect analysis were conducted. The simple main effect for sojourner role (Table 58) indicated a statistically

significant difference ($p = .04$, partial $\eta^2 = .09$) in language proficiency at post-sojourn between the two groups. Here, Single Honour students scored on average significantly higher than Joint Honour students. There was also a significant effect of time (Table 59) on language proficiency for both Joint Honour learners ($p = .001$, partial $\eta^2 = .33$) and Single Honour learners ($p = <.001$, partial $\eta^2 = .54$). The finding of the two-way mixed model ANOVA has indicated that sojourners return home significantly more proficient than when they departed, and that those who learn one language, tended to make more substantial progress than those who were learning two. Moreover, the results showed that sojourners, particularly Single Honour learners, tended to show stronger development in the second half of the year (see Table 59 and Figure 13).

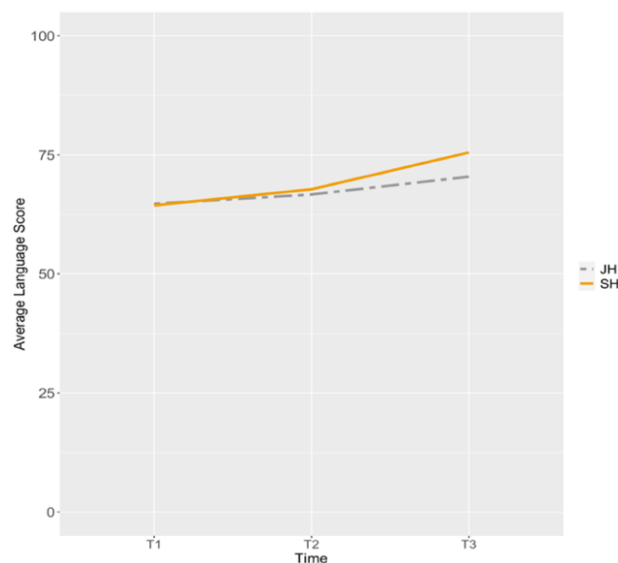
Table 58: Simple Main Effect for Group on Language Change

Time	Honours programme				Mean diff.*	F – statistic	Effect size (η^2)
	Joint Honours (n = 19)		Single Honour (n = 22)				
	M	SD	M	SD			
Pre	63.00	11.06	64.40	14.41	1.40	$F(1, 42) = .132, p = .71$.003
Mid	66.68	9.16	67.77	14.32	1.08	$F(1, 39) = .081, p = .77$.002
Post	69.09	10.62	75.50	10.22	6.40	$F(1, 42) = 4.308, p = .04^*$.093

Table 59: Simple Main Effect for Time on Language Change

Honours programme	T1 – T2			T2 – T3			T1 – T3			F - statistic	(η^2)
	M*	SE	p - value	M*	SE	p - value	M*	SE	p - value		
Joint Honours	1.94	1.65	.76	3.74	1.13	.01	5.68	1.23	.001	$F(2, 36) = 9.056, p = .001$.33
Single Honours	3.36	1.55	.12	7.73	1.69	.001	11.09	1.54	<.001	$F(2, 42) = 25.214, p = <.001$.54

Figure 13: Language Change Interaction Effect (Time * Honours Programme)



5.3.7.2 Perceived language changes

Alongside the C-test, learners were also asked to rate their perceived ability in the four language skills of speaking, writing, listening and reading, as operationalised by 12 ‘can-do’ statements. In line with earlier paired sample t-tests, perceived linguistic competencies changed was calculated by comparing scores post-sojourn to those of pre-sojourn. For those studying two languages, perceived scores were again averaged across both languages. Thirty language learners completed all datapoint necessary to conduct a paired sample t-test.

Table 60: Perceived Linguistic Ratings (Means/SDs)

Skill	T1		T3		Change Score		Paired-sample T-test	Cohen's <i>d</i>
	M	SD	M	SD	M	SD		
Listening	6.21	1.36	7.83	1.48	1.62	1.40	$t(29) = 6.358, p = <.001$	1.16
Reading	6.09	1.52	7.46	1.48	1.36	1.53	$t(29) = 4.870, p = <.001$.88
Speaking	6.02	1.58	7.49	1.52	1.47	1.52	$t(29) = 5.285, p = <.001$.96
Writing	5.67	1.69	7.16	1.43	1.40	1.56	$t(29) = 4.895, p = <.001$.89

The output (Table 60) of the paired sample t-tests showed language learners to perceive a significant improvement in all four skills. The table indicated learners to perceive the largest gain in listening ($m = 1.62, sd = 1.40$) followed by speaking ($m = 1.47, sd = 1.52$), writing ($m = 1.40, sd = 1.56$) and lastly reading ($m = 1.36, sd = 1.53$). The large standard deviations do however suggest substantial individual differences in the extent change was perceived. A two-way repeated-measures ANOVA indicated that no skill displayed greater change over time than another ($F(3, 87) = .362, p = .78$).

5.3.7.3 Qualitative findings

During the focus group discussions, learners were asked whether they perceived living abroad to facilitate their linguistic growth. Supporting the quantitative data described above, these learners generally perceived growth in the skills of listening and speaking. Participant 2 highlighted how for them, listening improved the most because it was the skill that was most frequently utilised:

“I would say that my listening improved the most. I think that was because I was studying, and the classes were in Italian, so I was forced to listen for an hour and a half at a time to someone speaking Italian.” (participant 2)

Participant 5 chose to highlight how living in a country (Peru) put them in situations where using the L2 was required, and a result believed their communicative ability had vastly improved:

“I feel like my Spanish definitely got better across the year abroad and likewise it was about the confidence. I find it more difficult to speak the L2 with native speakers, but abroad because I had to speak Spanish and there was no other alternative option it helped me build confidence and made speaking easier.”

(participant 5)

However, this degree of immersion was not generalisable across learners. Those who undertook their year abroad in westernised countries, where English was used as a Lingua Franca, continually felt frustrated by the lack of opportunities to use the L2 as highlighted by participants 1, 2 and 6. Whether the year abroad can be considered an *immersive* experience has been debated in the literature (Coleman, 2015; Jackson, 2018). For some learners in these discussions, the sense of immersion was missing, and this consequently limited the extent of perceived growth.

“I found it really frustrating in Austria because I was living in quite a touristy area and as such, they were able to speak English. It was really frustrating because I was there to speak German and so I would say that my German didn’t really improve that much.” (participant 1)

“My frustration is that I continued to find myself in settings where there was a lot of English. I thought that I would be more immersed in purely Italian settings and as much as I was in some respects, I wasn’t using my Italian as much as I wanted to be.” (participant 2)

“I think for me it was just not being able to speak French as much as I wanted to. I was teaching in a school and I was teaching English so that was at least 12 hours a week where I couldn’t practise French.” (participant 6)

Participant 9 was a Joint Honours language learner and chose to note their frustration at how they felt that their speaking ability was strengthened in one language and weakened in the other. Participant 11, on the other hand, was a Single Honour learner, finding the year abroad to be particularly beneficial to their linguistic development due to the opportunities afforded to them. Participant 11 also chose to highlight how their growth occurred after overcoming initial linguistic difficulties and indicates that a period of linguistic adjustment was required before acquisition could occur. This comment supports the notion of a threshold hypothesis (Cummins, 1976, DeKeyser, 2014) which posits that a certain proficiency *threshold* should be met before learners can take advantage of the opportunities afforded to them while abroad.

“I went away for 14 months in two countries. What I have found coming back is the country that you were most recently in, your language skills are better in the speaking classes. Whereas if you do two languages, your second language is really suffering now so it can feel like one step forward, two steps back sometimes.” (participant 9)

“I found [the year abroad] extremely helpful. I was in one country only so got a pretty concentrated period of development. I felt my language skills really improved in the second half of the year once I had picked up all the skills associated with conversation starters and got over the initial language hurdles.” (participant 11)

On the other hand, some individuals noted how they felt their writing and syntactic knowledge displayed limited progression. This lack of development was associated with limited opportunities to practice the L2 skill. Participant 10 noted how they continued to meet with informal language use (e.g., slang) and as language was predominantly used for communicative purposes. Likewise, the limited need to complete writing tasks while abroad meant that for some, such as participant 2, there was an overall sense of disappointment. These findings support the notion that individuals must remain highly disciplined on a year abroad in order to gain exposure to all four skills and make an effort to practice skills which are not immediately available to them.

“I have found that coming back I have forgotten a lot of the grammar. I know the simple stuff; the stuff you use every day and phrases you don’t really have to think about but having to think about more complex grammar stuff I have forgotten that, but I can say everyday language items such as slang proficiently.” (participant 10)

“With writing, apart from a few Italian lessons when I was studying, I didn’t do as much writing as I would have wanted to and so I don’t believe my writing developed as much as I wanted it to.” (participant 2)

In sum of the qualitative data, it has provided interesting insights into the perceived changes of language learners on a year abroad. These findings have generally supported those found in the quantitative analysis, indicating that in general, learners perceive themselves to become more proficient during the year abroad, particularly in the skills of listening and speaking.

5.3.7.4 Predictors of language change

The analysis above has indicated individual differences in overall proficiency gain, which warrant further investigation. As such, Section 5.3.7.3 examines whether the measured personality

variables can predict subsequent language gain, guided by the following research question, examining data collected from language learners of the sojourn sample:

RQ12: Does personality serve as a valid individual difference in linguistic gains made?

As noted by Dörnyei & Ryan (2015), personality has often been overlooked when evaluating why some language learners improve more than others. While the sample is small scale, the findings hope to provide insights into the importance of personality in predicting language change. Given the small sample, the regression models were limited in the number of variables which could be added. In light of the literature explored (e.g., Field, 2013), the maximum number of variables added to the final model could not exceed three.

Consistent with previous SLA literature (e.g., Beattie et al., 2014; Rees & Klapper, 2007), a change score (i.e., T3-T1) was calculated for each participant. Given that the research question referred explicitly to *personality*, only those variables associated with the *person* was tested. In line with earlier regression models, models were built in a stepwise fashion by inputting the learner's baseline L2 proficiency score first and by itself (Model 1). Secondly, Honours programme was added as a controlled factor (Model 2) as earlier ANOVA analysis indicated that this variable influenced language change and should, therefore, be controlled for. Lastly, the personality variables which reached significance in the preliminary analysis were added (Model 3).

Prior to building the full regression model, a series of preliminary regressions were conducted. In each regression, average language change (T3-T1) served as the dependent variable, while average baseline proficiency and the relevant trait variable served as the independent variables. Table 61 identifies the output of this analysis, which demonstrated that after controlling for baseline proficiency, only baseline openness (partial $r = .32$) significantly predicted language gain.

Table 61: Testing Potential Covariates of Language Change from T1 to T3

Variable Name	Definition	t-statistic and sig.
Baseline broad trait scores		
Openness	Trait openness as measured by the BFI	t (44) = 2.228, p = .03*
Conscientiousness	Trait conscientiousness as measured by the BFI	t (44) = .943, p = .35
Extraversion	Trait extraversion as measured by the BFI	t (44) = -1.233, p = .22
Agreeableness	Trait agreeableness as measured by the BFI	t (44) = .814, p = .42
Neuroticism	Trait neuroticism as measured by the BFI	t (44) = -.603, p = .55
Narrow traits		
Anxiety	Average anxiety score across the year	t (41) = -1.711, p = .31
Resilience	Average resilience score across the year	t (41) = 1.341, p = .18
Curiosity	Average curiosity score across the year	t (41) = 1.489, p = .14

Table 62 presents the output of the full regression model. All assumptions regarding linear regression (see Field, 2013) were met. The output indicates that baseline proficiency accounted for the majority of the variation in L2 proficiency gains, explaining 30.2%. The addition of Honours programme (Model 2, $p = .001$) and baseline openness (Model 3, $p = .03$), both significantly improved the prediction of L2 gain. Model 3 explained 51.9% of the variation in the learner's overall L2 change. The relationship between these variables and overall proficiency development will now be examined.

Table 62: Multiple Regression of Identified Covariates in L2 Proficiency Change (Baseline Openness)

Factors	Unstan. Coefficients		Standardised coefficients				
	B	SE	Beta	T	Sig.	zero-order	partial
Model 1							
Constant	8.59 (6.81, 10.36)	.088		9.758	<.001		
Baseline proficiency	-.298 (-.439, -.156)	.070	-.549	-4.259	<.001	-.549	-.549
Model 2							
Constant	11.30 (9.06, 13.54)	1.10		10.195	<.001		
Baseline proficiency	-.310 (-.436, -.184)	.062	-.571	-4.970	<.001	-.549	-.613
Language learner type	-5.43 (-8.60, -2.26)	1.57	-.398	-3.463	.001	-.366	-.476
Model 3							
Constant	11.17 (9.03, 13.32)	1.06		10.534	<.001		
Baseline proficiency	-.312 (-.433, -.192)	.059	-.576	-5.245	<.001	-.549	-.638
Honours programme	-5.17 (-8.21, -2.13)	1.50	-.378	-3.439	.001	-.366	-.478
Baseline openness	3.68 (.339, 7.03)	1.65	.244	2.226	.032	.262	.332

Model 1: $R^2 = .302$, Adjusted $R^2 = .285$, $F(1, 42) = 18.138$, $p = <.001$; Model 2: $R^2 = .460$, Adjusted $R^2 = .433$, change in $R^2 = .158$, change in $F(1, 41) = 11.993$, $p = .001$; Model 3: $R^2 = .519$, Adjusted $R^2 = .483$, change in $R^2 = .060$, change in $F(1, 40) = 4.954$, $p = .032$

Baseline proficiency

Baseline proficiency was the strongest predictor in Model 3 ($t(44) = -5.24, p = <.001$), explaining 30.2% of the variance in overall proficiency change. The negative partial correlation (partial $r = -.63$), indicated that individuals who began the year abroad with higher proficiency tended to make less gain than those who were of lower ability. This association may be reflective of a methodological artefact, whereby pre-scores naturally show an inverse relationship with a change score when scores are bounded. This observation has been repeatedly found in the literature (Collentine, 2009; Dewey et al., 2012) and have been explained by the presence of ceiling effects.

Honours Programme

The addition of Honours programme (Model 2) accounted for a further 15.8% in the variation of proficiency change scores. As Honours programme was dichotomously coded (0 = Single Honours/1 = Joint Honours), the regression model estimated Single Honour language learners to demonstrate a greater gain of approximately five marks than their Joint Honour peers.

Baseline openness

Adding information on baseline openness made a significant contribution to the prediction of L2 development ($t(44) = 2.226, p = .03$). The association between baseline openness and L2 development was positive and small to moderate in size (partial $r = .33$), indicating that those students who perceived themselves more open at baseline tended to make greater L2 gains. This relationship is theoretically supported because those who are more open will likely have a stronger desire to seek out L2 orientated experiences (Dewaele, 2013).

5.3.7.5 Summary of findings

Section 5.3.7 has looked to examine whether a) learners return home more proficient in their L2 than whence they departed and b) to ascertain whether personality variables accounted for differences seen in the rate of change.

The findings have indicated that on average, a language learner improved by 8.59 marks in their chosen L2, which represented a significant improvement from baseline. However, as Joint Honour and Single Honour language learners were accounted for in this study, further analysis

was warranted regarding whether sojourning was of more benefit to those who studied only one language. Here, the results indicated that those who learnt one language benefited significantly more from the year abroad experience than those who learnt two languages. The average Single Honour student improved almost twice more than their Joint Honour peers.

Concerning the role of personality, this study has found that after controlling for baseline proficiency score, together with that of Language Honour programme, language learners who were more open at baseline tended to improve significantly more than those less open at baseline. While no other personality variable reached significance, this finding does tentatively indicate that personality can influence language gain abroad and that personality should be seen as a valid individual difference in the literature which warrants further investigation.

5.4 Discussion and conclusions

From an applied perspective, Study 2 has looked to examine whether the goals and objectives of a year abroad are realistically achieved. In doing so, it has looked to examine whether those who undertake a year abroad experience greater change in personality and well-being than those who remain at-home. Study 2 also investigated whether language students returned home significantly more competent in their linguistic ability and ascertained potential predictors which could account for differing rates of change. From a theoretical perspective, the study ascertained the breakdown of between-and-within-person variability in participants across all three levels of personality examined (i.e., broad trait, narrow trait, & state). It further identified whether within-person variability was systematically related to that of situational cues and characteristics.

To achieve the study's aims, a series of online questionnaires were disseminated within a repeated-measures design at differing intensities (ranging between 4 questionnaires and 84 questionnaires over the year). In order to provide further insights not immediately possible from the questionnaire data, a series of focus groups were held post-sojourn.

An important finding in Study 2 has been that sojourners became significantly more agreeable and curious over time compared to those who remained at-home. Regarding agreeableness, this finding supports previous literature (Niehoff et al., 2007; Zimmermann & Neyer, 2013) which explored broad trait change in L1-speaking German sojourners. Both studies found sojourning to lead to an increased rate of agreeableness change compared to the comparison group.

Curiosity has, to the best of my knowledge, not yet been explored using quantitative approaches, but nonetheless, findings of Study 2 support the published qualitative evidence (e.g., Gu & Maley, 2008; Tracy-Ventura et al., 2016) which has tended to find sojourners to perceive themselves more open and curious towards to new experiences and cultures post-sojourn. These findings are discussed more in section 6.2.1 of the discussion.

Regarding well-being, given the current focus on an at-home undergraduate sample (e.g., Bewick et al., 2010; Denovan & Macaskill, 2017), Study 2 has provided particular insights into well-being change and mechanisms of change within a sojourning sample. Here sojourning was found not to result in significantly higher well-being over time compared to remaining at-home, indicating sojourning had little facilitative benefit to the well-being of the average individual. This has provided an important insight for to date, little is understood regarding the extent to which sojourning fosters positive well-being (Lees, 2020; Potter, 2020), particularly within a longitudinal, repeated measures design.

Linguistically, language learners were found to return home significantly more proficient in their chosen L2 than when they departed for the year abroad. This conclusion, however, belies the extent to which the year abroad benefits Single Honour language learners compared to Joint Honour language learners. A learner learning a single language was significantly more proficient in that language than a Joint Honour learner was in any of their two languages. This is an important insight and one which has yet to be explored in the literature, which have focused primarily on Single Honour learners (Hessel, 2016; Mitchell et al., 2017a). Where Joint Honour language learners were present (e.g., Rees & Klapper, 2007), this distinction has not been discussed. Moreover, baseline openness scores were found to be a significant predictor of language change abroad, indicating those who were more open at the beginning of the year abroad became significantly more proficient than those less open. Previous literature (Ożańska-Ponikwia & Dewaele, 2012) has found open learners to be more expressive and willing to communicate and consequently assuming these learners will then make significant improvement, as per SLA theory (Swain, 1985). This finding tentatively (given the sample size) confirms these assumptions and indicates that personality can influence language change and personality should be seen as a valid individual difference in the SLA literature.

The findings of the state personality component indicated that the more intensive the repeated measurement, the more of the total variability that is accounted for by within-variability. That is to say, state personality demonstrated the most amount of within-person variability (i.e., fluctuations around one's own mean), while broad traits demonstrated the least amount of within-person variability. Given that broad traits are known to be stable over time, this finding is in line with current theory. Importantly, however, findings have shown that within-person variability, at least for agreeableness, was systematically associated with the situational characteristic of Duty, with, *sojourners* perceiving themselves to be significantly more agreeable when Duty was higher compared to *non-sojourners*. This lends support to the notion of situation contingencies (Fleeson 2001; 2007) and supports the perspective that within-person variability warrants further investigation as opposed to being considered purely error variance.

A limitation of the state personality component, however, was the presence of missing data, particularly within the sojourn sample. For many sojourners, this was borne out of not having the internet to complete the questionnaire. Open WIFI access is not available in many of the countries visited in this study, and consequently, some individuals failed to complete measures when not at-home. This links succinctly to the next limitation, whereby much of the collected data is representative of when the participant was at-home (i.e., sojourners' residences when abroad) because this is where they had the time and internet access to complete the measure. Therefore, the findings are more reflective of behaviours when at-home, rather than in any other location and limit the generalisability of the findings to other more foreign environments. Lastly, unbeknownst to the author, the random weeks selected fell on holidays across European countries. Consequently, it was felt by some individuals that their state personality measures were not always realistic of their normal behaviours as they were not in their normal environments. Instead of studying or working, for example, individuals were travelling or on holiday, where emotions may be reflective of more positive affect than their normal behaviours.

A further limitation of the study as a whole has been that analysis is limited to a series of variables which were chosen based on subjective and literature-based reasoning. It cannot be ruled out, however, that potentially relevant variables were not included in the regression models (e.g., attitudes towards host country). Nonetheless, as Dewaele (2013) notes, no study could ever fully account for potentially important variables, and as such, any choice could be labelled as *reductionist*. Moreover, the relatively small sample size impeded the number of variables which

could be tested simultaneously in the final models, hence the need to first conduct preliminary regression analysis. A preferred method would have been to account for all variables (i.e., *person, environmental, & program characteristics*) in one model so that all variables could be controlled for and held constant.

The findings of Study 2 can be summarised as follows:

- Sojourners became significantly more agreeable and curious over time compared to those who remained at-home.
- Within-person variability should not be treated as error variance. This variability was shown to be systematic with perceived changes in the environment, and average state Agreeableness was found to predict several outcomes at the between-person level.
- Sojourning did not appear to have a multiplicative effect on psychological well-being. Nonetheless, sojourners, who felt part of a community and partook in extra-curricular had on average higher well-being across time.
- Language learners returned home significantly more proficient than at baseline, with those who learnt one language returning home more proficient than those learning two. Baseline openness was found to be a significant predictor of language gain.

This chapter has served as Study 2 of the thesis and has collected data on a series of outcomes, mainly that of personality, well-being, and language. The final chapter looks to discuss these findings in more detail and place them within the findings of study 1 also, as well as the overall objectives of the thesis.

Chapter 6: Discussion and Conclusions

6.1 Returning to the Purpose of the Thesis

An ever-growing number of learners undertake a study abroad experience each year. Serving as a natural hiatus to academic study, it can provide an opportunity for linguistic and cultural immersion. Given the level of resources required in planning and undertaking a study abroad for stakeholders and students alike, it is perhaps surprising that not more quantitative research has taken an evaluative approach in ascertaining whether the experience achieves its purpose. Consequently, the tradition of viewing a sojourn as a *transformative* life experience in relation to personal and linguistic competencies continues to remain constant (Johnson & McKinnon, 2018). This view, must, however, as noted by Bryan (2008) over a decade ago, be consistently empirically challenged, and it is this that this thesis has sought to achieve.

Ascertaining whether a sojourn experience meets its purpose is of vital importance when managing student expectations, which for British students, in particular, may be particularly heightened, according to Johnson and McKinnon (2018). This is because the year abroad experience has been commercialised in the manner it is 'sold' using glossy images and idealised testimonies. Moreover, British-domiciled students tend to view education through consumerist terms. While tuition fees are high (although capped while abroad), students tend to be offered a high level of learning and pastoral support. Many students, therefore, expect the same when abroad but are left disappointed if they feel the support offered is not *value for money*.

The thesis has aimed to challenge and evaluate the role of sojourning in fostering the skills and behaviours in three domains: that of personality, language and psychological well-being.

Study 1 consisted of a systematic review, with the purpose of synthesising the large body of literature which, to date, has explored the development of linguistic competencies on a study abroad. By using inclusion/exclusion criteria, together with quality appraising included studies, the review aimed at identifying causal inferences regarding linguistic gain as a direct effect of the sojourn experience.

In Study 2, data were collected longitudinally on a range of outcome measures (e.g., broad trait personality) within a quasi-experimental design, whereby individuals who were abroad served as the intervention group and those at-home served as the comparison group. Due to the inclusion of a comparison group formed through self-selection, Study 2 has assessed, in a limited way, whether sojourning fosters an accelerated change in personal growth and well-being compared to remaining at-home.

It is in the construct of linguistic competencies where Study 1 and Study 2 complement each other. Study 1 enabled limited causal inference to be drawn on the effectiveness of the study abroad. Given the lack of possible allocated control group for the linguistic component in Study 2, observation of a direct effect was not possible, and instead Study 2 provided insights into the potential predictors of linguistic gain during a year abroad experience.

Before integrating findings with previously known literature, the aims and objectives, together with findings, must be placed in the context of study design. Given that the overarching research question is evaluative in nature, the study design implemented must allow for causal inference. From a methodological standpoint, RCTs can be considered the most robust study design when looking to attribute an effect to a particular intervention, and as such warrant the strongest causal inference (Bryman, 2012). However, given the moral and ethical issues surrounding randomly selecting participants to experience a study abroad, an RCT design is near inoperable within the study abroad context. Although less robust than that of an RCT for they lack random assignment, QEDs can still allow for causal inference, and it is this type of design, both the systematic review and Study 2 has looked to utilise.

In analyses where a control group has been included, the strength of determining causal inference has been increased by collecting and controlling for baseline scores. This allows for analyses/findings to take into account just how different the two groups were at the start of the study and without this, the study design would be considered *weak* (Christensen et al. 2014). In brief, baseline scores indicated participants in each learning context to be similar in all outcome variables apart from that of neuroticism. Nonetheless, I acknowledge the flaws when attempting to determine causal inference when undertaking a QED, but in light of the topic under study, it was the most robust design which could be implemented when answering an evaluative-type research question.

6.2 Does the Year Abroad Work? Evaluating the Value of a Sojourn Experience

6.2.1 Evaluating the value of sojourning on personal growth (Study 2)

Study 2 conceptualised personal growth through means of broad and narrow traits. Regarding the broad traits first, only agreeableness demonstrated a significant interaction term, suggesting that *sojourners* became significantly more agreeable over time than *non-sojourners*. This significant interaction effect remained when controlling for pre-existing (i.e., baseline broad trait scores) and time-interval (i.e., narrow traits) variables. Consequently, the results imply that the observed differences in the rate of agreeableness change can, to some extent, be causally attributed, to the learning context. As such, it may be argued that sojourning does foster personal growth, at least in behavioural tendencies associated with agreeableness such as trustworthiness and kindness.

This finding is in line with the few studies which have explored broad trait personality change on a study abroad through implementing a quasi-experimental design. Zimmermann and Neyer (2013) studied 1,134 L1-speaking German undergraduates, 527 of whom went abroad, and 607 of whom remained at-home. Using multivariate latent change models, the study captured broad trait scores at three timepoints during the academic year. The results indicated that at each timepoint, sojourners scored $\beta = 0.07$ units higher than non-sojourners in agreeableness, with Zimmerman and Neyer (2013, p. 525) describing sojourners as experiencing an “accentuated increase of agreeableness” compared to those who remained at-home. Similarly, Niehoff et al. (2017) explored broad trait change over one semester in 221 L1-speaking German undergraduates, 93 of whom studied abroad, and 128 of whom remained at-home. Using multivariate regression models, their findings indicated that after controlling for learning context and baseline scores, a direct effect of sojourning was found for agreeableness only.

The question subsequently arises as to why sojourning influences the extent of change in trait agreeableness in particular. As a trait, agreeableness is associated with characteristics such as being friendly, helpful and unselfish, with agreeable individuals tending to be interpersonally altruistic and co-operative (Soto, John, Gosling, & Potter, 2008). Integration is possibly accelerated if members of the host community approach sojourners, which is more likely when sojourners demonstrate friendly behaviours (Greischel et al., 2016). Moreover, by demonstrating pro-social behaviours, tasks such as becoming acquainted with the local environment and

learning from host community members are facilitated (Greischel et al., 2016). In both long-term expatriate studies (e.g., Bakker, van Oudenhoven, & van der Zee, 2004; Huang, Chi, & Lawler, 2005) and sojourning studies (e.g., Ward, Leong, & Low, 2004; Geeraert et al., 2019), agreeableness has been positively related with that of socio-cultural adjustment. Consequently, by displaying more agreeable trait-like tendencies over time, individuals are improving their chances of being accepted by the host community and fostering a sense of belonging.

For those at-home, however, support and social networks are likely already in place, and learners are, as such, more likely to already feel part of the home community. As a result, the need to demonstrate more agreeable trait-like tendencies is lessened. For *non-sojourners*, previous literature (e.g., Parker, Lüdtke, Trautwein, & Roberts, 2012; Schurer et al., 2015) had tended to show that the transition between post-secondary and tertiary education results in substantial agreeableness change after which, the extent of agreeable change slows down.

Furthermore, in line with Zimmermann and Neyer (2013) and Niehoff et al. (2017), the findings of Study 2, indicated that at the between-group level, sojourning had little effect of conscientiousness change. This finding may, however, belie the importance of sample make-up. Both Zimmermann and Neyer (2013) and Niehoff et al. (2017) investigated changes in learners who were in full-time study while abroad, yet the findings of section 5.3.4.2 indicated that employment status while abroad significantly impacted the extent of conscientiousness change observed. Here, it was found that those who were in employment during the year abroad (i.e., worked or taught) appeared to experience more significant conscientiousness change than those who remained in full-time study. While this finding is in line with past research (Lüdtke et al., 2011), it is the first time this change has been explored within a sojourning context. Consequently, further insights could be gained by analysing interaction effects between those who were in employment during the year abroad and those who remained at-home. It may be hypothesised that study abroad can have a direct effect on conscientiousness change, but only when individuals are in employment. While this distinction was not explored, it provides an avenue for further research.

However, Zimmermann and Neyer (2013) also found a direct effect of sojourning on openness ($\beta = 0.05, p = .03$) and neuroticism ($\beta = -0.13, p = <.001$), findings of which were not duplicated in Study 2. Given that openness was close to the boundary of significance ($\beta = 0.03, p = .07$),

were a larger sample size employed, it is possible significance would have been found. As the sample size in Study 2 was relatively small, the threat of encountering a Type II error was increased. Regarding neuroticism, it was perhaps surprising that sojourners in Study 2 did not show more accentuated change compared to those who remained at-home ($\beta = -0.03, p = >.05$). The difference in findings may be because roughly half of the sojourners in Study 2 changed country at the mid-point of their year abroad, whereas those in Zimmermann and Neyer (2013) remained in the same country for the entire duration. Consequently, a subset of learners in Study 2 once again experienced the psychological agitation met by moving to a different country, which may have resulted in a higher group average neuroticism score at T2 and T3. Perhaps a fairer comparison would be to detect an interaction effect between those at-home and those who remained in the same country the whole year.

Turning attention to the narrow traits, the findings of Study 2, have provided original insights into the impact of sojourning on the development of anxiety, curiosity, and resilience, behaviours which have to date been overlooked through quantitative approaches. The most pertinent finding has been that of curiosity, whereby a significant interaction between learning context and time was found. This finding is generally supported by the literature (e.g., Hadis, 2005; Tracy-Ventura et al., 2016; Gu & Marley, 2008; MacLeod & Wainwright, 2009; Williams, 2015), which through past testimonies (e.g., focus groups & interviews) have reported that sojourners return home more open-minded and accepting of foreign cultures and ideas. Less conclusive evidence was found for anxiety and resilience. While a significant interaction effect for anxiety existed, suggesting sojourners became significantly less anxious over time compared to non-sojourners, this interaction was non-significant after controlling for resilience. Moreover, it cannot be discounted that the observed differences in anxiety scores over time were not caused by the different pressures faced by each group. In section 5.2.1.4, it was noted that 3rd year at-home undergraduates face a number of pressures not faced by sojourners (e.g., exam pressure; job-hunting etc.), which may potentially exacerbate anxiety in the at-home context over that of being abroad. An alternative approach would have been to explore changes in a 3rd year abroad sample and a 2nd year at-home sample. However, it was decided that this would not be suitable as the two groups of learners may differ substantially on personality variables at baseline given the relationship between age and maturity. Furthermore, given that sojourners were found to become significantly less anxious over time (i.e., main effect of time), this finding does indicate that sojourning could have, at least to some degree, an effect on decreasing anxiety. No

significant interaction term was found for resilience, suggesting the average sojourner did not become significantly more resilient over time compared to remaining at-home. These findings go against the general assumption that sojourning should facilitate these traits and as such warrant further attention with a larger sample size.

In sum of RQ2, given the study design, it can be tentatively suggested that sojourning does foster personal growth for the *average* individual in succinct and specific traits. Of further interest was understanding the directionality and the extent to which sojourners changed regarding their personality (RQ3), together with understanding the possible mechanisms of this change (RQ4).

Study 2 found a main effect of time for neuroticism and openness, indicating sojourners to become significantly more emotionally stable and open over time. Regarding emotional stability, these findings support those of Tracy-Ventura et al. (2016), who captured personality change in 58 British sojourners on a year abroad using the MPQ. By using paired sample t-tests, Tracy-Ventura et al. (2016, p. 122) demonstrated sojourners to return home significantly more emotionally stable ($p = .04$) concluding that “RA [Residency Abroad] is an example of a life event with the potential to influence personality change.” However, the strength of this conclusion is somewhat contradicted by the findings of Study 2, for while the main effect of time is significant, the interaction term (i.e., learning context * time) was non-significant, suggesting that this increase in emotional stability was not as a direct result of the sojourning experience itself. It may as such be argued that a confounding or unknown variable is influencing this change and weakens the evidence regarding the extent to which living abroad can be causally attributed to the observed depreciation in neuroticism.

Conversely, the sample size may have been too small in the current study (i.e., Study 2) to detect a significant interaction effect (i.e., Type II error). When looking at the two groups independently, those at-home appeared to demonstrate little change in emotional stability ($\beta = -.001$), while for those abroad the trajectory was steeper ($\beta = -.003$) offering a tentative conclusion that the learning context may at least influence neuroticism change. Moreover, previous literature has also found neurotic trait-like tendencies to generally decrease over time (e.g., Zimmermann & Neyer, 2013; Savicki & Price, 2017), with the observed finding being attributed to the fact that anxiety is highest pre-sojourn due to learners entering *the unknown*. As learners acclimatise and integrate into the host community, these neurotic trait-like tendencies tend to dissipate.

Regarding openness, evidence concerning the main effect of time has been mixed. Tracy-Ventura et al. (2016) and Niehoff et al. (2017) found no evidence to indicate trait-like tendencies associated with openness to change significantly over time in sojourners. Schartner (2016) found sojourners to score significantly lower at post-test in open-mindedness compared to at pre-test conducted ten months earlier in 143 postgraduate international students. Schartner (2016) explained this finding by noting sojourners were already high in openness at pre-test (i.e., regression to the mean). On the other hand, Zimmermann and Neyer (2013) found sojourners to become increasingly more open over time, finding that learning context was a causal mechanism in accounting for this change.

When examining individual trajectories in both the broad and narrow traits (RQ3), substantial individual differences existed. While some individuals strengthened in their trait-like behaviours, others weakened. This heterogeneity corroborates with the view of Coleman (2015, p. 37), who notes that “referring to ‘the’ study abroad experience is a patent absurdity”. This insight is important and continues to challenge the view that study abroad has universal benefits for all (e.g., the British Academy, 2011). This finding also raises the question of *why do some individuals experience greater change than others while abroad?* To date, little evidence has been forthcoming as to the possible antecedents of personality change with the few studies which have done so focusing solely on relationship fluctuation (Zimmermann & Neyer, 2013; Gresichel et al., 2016). Study 2 has taken a broad approach to variable selection based on both literature and subjective opinion and in doing so, has examined variables which hitherto have not been explored.

Concerning RQ4, the regression analyses indicated that *loneliness* appeared to inhibit change in openness, conscientiousness, extraversion and agreeableness and led to an accentuated increase in neuroticism. Conversely, *participation in extracurricular activities* appeared to foster broad trait growth in openness, conscientiousness, extraversion and agreeableness, while it appeared to inhibit neuroticism.

The relationship between loneliness and personality has been explored more generally at a cross-sectional level (e.g., Vanhalst et al., 2012; Teppers et al., 2013), with elevated levels of loneliness being correlated with higher levels of neuroticism and lower levels of extraversion and agreeableness. While less is known regarding the longitudinal relationship between loneliness and

personality change (Stephan, Sutin, & Terraciano, 2014), persistent loneliness is likely to manifest itself into depression and anxiety. Over time, this consistent behaviour may become integrated into an individual's psychological system, where it leads to enduring changes in the broad traits (Soto, 2015). The effect of loneliness on an individual's personality may be, even more, exacerbated in a sojourning context, given that the familiar support networks are no longer available. Moreover, the behavioural risk factors which are associated with loneliness (e.g., physical inactivity; smoking) may also influence personality trait development (Mund & Neyer, 2019). Conversely, participation in extracurricular activities is likely to facilitate integration into a host community and increase the rate in which acculturation takes place (Shiner & DeYoung, 2013). This integration is of vital importance in fostering opportunities for intercultural communication and enabling learners to seek out new and exciting opportunities.

Previous literature (e.g., Zimmermann & Neyer, 2013; Ward & Chang, 1997; Ward et al., 2004) have challenged future research to employ a more heterogeneous sample in order to ascertain whether programme characteristics influence the extent to which personality develops abroad. This current study has explored several predictors associated with sample make-up, including host country, type of sojourn programme (compulsory/voluntary) and whether the year abroad was split. The findings indicated that no programme variable other than that of sojourner role (in conscientiousness change) appeared to predict broad personality trait change. As such, an argument can be made that researching a homogeneous sample can still offer important insights and hold value in the literature.

The qualitative findings have also highlighted constructs more akin to those identified in the intercultural literature (see section 2.10). For example, individuals spoke of being better able to overcome adversity (cross-cultural adaptability), finding a motivation and desire to seek out new interactions (intercultural communication apprehension), and a general enjoyment in partaking in new experiences and learning from other culture (intercultural competence). These findings indicate that a period abroad can be of value in developing the intercultural competences of those who undertake it. It should, however, be noted that a few individuals spoke of experiencing no change and again re-iterates the importance of appreciating the heterogeneous nature of perceived change by sojourners and that it may not be of intercultural value to all. While Study 2 investigated personality from a psychological perspective employing the widely accepted Big Five framework, the qualitative findings also support perspectives identified within the interculturality literature.

6.2.2 Evaluating the value of sojourning on psychological well-being (Study 2)

Study 2 also looked to capture changes in psychological well-being, which is becoming an ever-important area of discussion in both the sojourning and at-home learning contexts. In Study 2, greater attention was given to the sojourning context because, to date, well-being change and its mechanisms have been explored more within the at-home context (e.g., Bewick et al., 2010). Moreover, when exploring sojourners, well-being has typically been subsumed within the term *psychological adjustment* (e.g., Ward & Searle, 1991) and often examined retrospectively, through qualitative measures (e.g., Lees, 2020; Potter, 2020) as opposed to longitudinal means.

The results of RQ7 provided no evidence to indicate sojourning resulted in greater positive well-being over time compared to remaining at-home. Regarding the variable time, the group average well-being score was significantly higher in the final few months than at baseline. This finding concurs with previous literature (e.g., Walker & Braskamp, 2008; O'Reilly, Ryan, & Hickey, 2010), which have undertaken a pre-post-test design, concluding sojourning to foster positive well-being. Study 2, however, also took a repeated measurement of well-being on a monthly basis (RQ8). The main effect of time proved non-significant, indicating sojourners did not experience higher well-being over time while abroad. The question, therefore, arises as to why well-being is significantly higher when many are preparing to leave the year abroad or have left already compared to at baseline. The general linear trajectory, as presented in Figure 10, would indicate that the average sojourner experienced small incremental increases in well-being each month as opposed to a general relief that the experience is over. Such an increase may be brought on by the assumed continual improvement in host relations as individuals integrate and adapt to the new culture.

These nomothetic approaches, as undertaken by the aforementioned studies, also often ignore the within-group variation in well-being trajectories over time. Indeed, the results in Study 2 demonstrated significant within-group differences with some perceiving increased well-being over time, and others experiencing a depreciation in their well-being, with this finding in line with broad and narrow trait analyses.

This variability warrants further investigation as has been attempted in Study 2, for studies (e.g., Sam, 2001; Walker & Braskamp, 2008; Potter, 2020) have generally taken a top-down approach in understanding mechanisms of positive well-being in a sojourning context. In doing so, well-

being tends to be viewed as a tangible construct which can be influenced directly by observable constructs. Study 2 has, on the other hand, looked to view these mechanisms using a bottom-up approach (RQ9), made possible by simultaneously capturing fluctuations in the narrow traits (i.e., anxiety, curiosity and resilience). The findings have indicated that through promoting stronger behavioural trait-like tendencies in these narrow traits, well-being will be naturally higher. As such, it is argued institutions would be advised in focusing on such trait-like tendencies and avoiding attempts to manipulate well-being directly.

Of interest to Study 2, was also ascertaining the association between *person*, *environmental* and *program* characteristics in relation to overall well-being (RQ10). The findings indicated that both the baseline *person* variables (conscientiousness and neuroticism) and *environmental* factors (loneliness, negative events, belongingness and participation) all predicted average well-being in sojourners.

Conscientiousness has traditionally been known for its strong association with psychological well-being. After all, conscientious individuals tend to believe that they shape and alter their environment in positive ways, solve problems which may arise and have a higher level of motivation towards achieving success (Carter, Guan, Maples, Williamson, & Miller, 2016). Similarly, neuroticism has long been associated with poor psychological well-being (e.g., DeNeve & Cooper, 1998; Argyle & Lu, 1990). Bardi and Ryff (2007), explored the relationship between psychological well-being and neuroticism after relocation (not associated with international mobility). The authors found neuroticism to be negatively associated with post-move autonomy, personal growth, positive relations and purpose in life. It can be said that Study 2 findings support the general literature, and provides generalisability to study abroad, a learning context, which to date has been largely understudied.

Loneliness is to date, the most researched of the variables accounted for in the current study, (e.g., Chavajay & Skowronek, 2008; Wawera & McCamley, 2019; Gu & Maley, 2008), and has typically been found to inhibit well-being, as evidenced by the findings of Study 2. Loneliness has long thought to be a stressor to an individual's well-being because it induces and exacerbate learner anxiety. When abroad, it may be posited that the impact loneliness has on well-being is exacerbated, because the support networks available at-home are no longer available.

The onset of negative events was found to be significantly and negatively associated with overall well-being. This finding contradicts Potter (2020), who found retrospectively, that learners perceived encountering, and subsequently overcoming, negative experiences to have an enhanceive effect on a learner's well-being. Potter (2020, p. 23) notes that this improvement is in line with the notion of eudaimonic well-being, focused on "individual striving, optimal functioning, and individual fulfilment." The difference in findings may be related to the time component of the two studies. Whereas Potter (2020) retrospectively assessed the relationship, this current study explored the relationship almost concurrently with the negative event itself. As such, the negative event, or the repercussions from it, may have still been ongoing when the measure was completed, inhibiting perceived well-being.

Conversely, belongingness in the host community has been demonstrated to mitigate the effects of loneliness and homesickness and as such, assumed to facilitate positive well-being (Holloway & Brown, 2008). Moreover, this connectedness can create a supportive learning environment, further facilitating a sense of purpose and satisfaction (Petersen, Divitini, & Chabert, 2008).

Similarly, those who partook in extracurricular activities were found to have, on average, significantly higher perceived well-being. Participation in these is likely to facilitate integration, for learners can acculturate faster and build stronger support networks. Moreover, this participation in extracurricular activities has repeatedly been found to have a positive effect on well-being in individuals of all ages (Oberle et al., 2019; Guilmette, Mulvihill, Villemaire-Krajden, & Barker, 2019), and as such extends these findings into the study abroad learning context.

6.2.3 Evaluating the value of sojourning on language learning (Study 1 & Study 2)

Improving linguistic competencies has consistently been considered an important outcome of a study abroad experience (Freed, 1995; Pérez-Vidal, 2014), and both studies in the thesis have, in some form, examined the extent to which linguistic proficiency changes during a sojourn. It must, however, be noted that due to the impracticalities of forming a comparison group in Study 2, only Study 1 can evaluate the role of learning context in accounting for the observed proficiency change, and it is here, where the discussion shall begin.

Study abroad has long been thought of as the superior learning context in which through linguistic immersion, language acquisition will be accelerated compared to remaining at-home (Llanes & Muñoz, 2013). Using a systematic review approach, Study 1 aimed to synthesise the

relevant literature in order to ascertain whether such an assumption holds true (RQ1). From including 1,533 studies in the first stage of screening, the review then identified seven studies which met the inclusion/exclusion criteria; and which utilised a study design which warranted limited causal inference. These seven studies focused on a range of linguistic outcomes and sampled students from both North America and Europe. The synthesis reported in Study 1 builds on the small pool of published reviews available: that of a meta-analysis conducted by Yang (2016) and a scoping review conducted by Tullock and Ortega (2017). This review is, to the best of my knowledge, the first to utilise the SWiM guidelines (Campbell et al., 2020) for reporting a systematic review which does not undertake a meta-analysis.

Given the recent influx in European-orientated literature (e.g., Mitchell et al., 2017a; Howard et al., 2020), the conducting of this review is timely. For example, Yang (2016) found 11 studies which met the inclusion/exclusion criteria when assessing the efficacy of sojourning on language change. Of these, only one investigated a European-domiciled sample, and all were published before 2012. Consequently, it was felt that a more up-to-date review would be of value.

Nonetheless, this review has encountered similar issues, like those which came before it. Like in Tullock and Ortega's review (2017), a number of studies were dropped, for despite having a comparison group, either failed to test for baseline equivalence or not report it. Similarly, as in the review by Yang (2016), the number of studies included in the final review ($n = 7$) is a small sample. This is possibly indicative of the general issues highlighted in the field regarding the formation of a suitable at-home comparison group when looking to infer causality (e.g., Meara, 1994; Freed, 1995; López-Serrano, 2010; Kinginger, 2009; Grey, 2018; Marijuan & Sanz, 2018). Nevertheless, the systematic review should be considered a robust appraisal of whether a study abroad facilitates the acquisition of a second language. The methodology of the review is open and transparent, both features often missing from published narrative reviews. Determining causal inference has been strengthened by ensuring baseline equivalence is achieved on the outcome measure (i.e., baseline linguistic proficiency) and providing an in-depth review of published studies which undertook relatively robust study designs for determining causality (i.e., quasi-experimental designs). As such, the conclusions presented, although formed from a small evidence base, can provide for refined insights in determining how effective sojourning can be in facilitating linguistic growth, and perhaps most importantly, ascribe this change to the sojourning context itself.

Furthermore, while the originality of Study 1 is predominantly in its methodology, the review has provided particular insights into the facilitative nature of the ERASMUS programme, something which is important for two reasons. Firstly, ERASMUS orientated literature has tended to be published in the past decade and as such narrative reviews dated prior to 2010 have not been able to explore the topic in any great detail. Secondly, ERASMUS participants are often demographically much different to their American counterparts (e.g., higher proficiency; differing living arrangements), all of which can influence findings. As such, findings of narrative reviews may be of less relevance to European readers. However, it is hoped that this systematic review (Study 1) can provide insights which are particularly relevant to European readers given its focus on published ERASMUS research.

Upon quality appraising the studies, those studies of the highest quality indicated sojourning to appear to have a multiplicative effect on oral fluency and general proficiency. This conclusion supports published narrative reviews on the topic (e.g., Llanes, 2011; Borrás & Llanes, 2019), but given the systematic review approach, these conclusions should be considered more robust and less biased than those found in previous narrative reviews.

Turning attention to Study 2, global proficiency was measured via a set of C-tests, where it was found that over the academic year, the average language improved by roughly 8.59 marks in their chosen language (RQ11). To put this comparatively, in the current study, sojourners improved by 8.59%. Using C-test also to measure global proficiency, Coleman (1996) reported gain scores of 16.36%, although this change was over two years, one of which was spent abroad. Rees and Klapper found the average L2 German learner to improve by 9.47% over the academic year, while Hessel (2016) found long-term sojourners to increase on average by 13.40%.

Study 2 has provided two important insights, which, to date, have often been overlooked in the SLA literature. These are a) the extent to which dual language learners are possibly disadvantaged during a study abroad and b) the role of personality in predicting linguistic gain.

Typically, language studies have explored Single Honour language students during a period abroad. This has been evidenced by the results of the mapping table in Study 1 (see Appendix D) in which, all included studies explored language change in a single L1, most frequently English. While this may be representative of language learners at large, in an L1 English speaking

environment, language learners tend to be dual language learners (e.g., French and Spanish), otherwise known as Joint Honour Language learners (Johnson & McKinnon, 2018). These learners typically divide their year abroad between countries which speak their learnt languages. As such, instead of spending up to 14 months in a single country, these learners spend up to 7 months in each. The impact of this change is yet to be explored in the literature (Winke, Gass, & Heidrich, 2019).

The findings of Study 2 found that Joint Honour language learners experienced less proficiency gain than Single Honour learners. This finding may be explained in accordance with SLA theory, which would posit that language acquisition is facilitated through direct exposure to the L2 (e.g., Swain, 1985; Krashen, 1985). Theoretically, those learning only one language would receive double the exposure compared with those learning two languages and the mean score change in each group would support this notion, with Single Honour learners demonstrating a mean increase of 11.09 marks compared to the observed 5.68 marks increase in Joint Honour language learners. As marks for each learnt language was combined for dual language learners, the observed limited growth in the second half of the academic year may be reflective of depreciation in L2 ability of the first host country. In the focus group discussion, one participant noted how they felt their L2 proficiency of their first host country regress in the 2nd half of the year, describing their proficiency change as *“one step forward, two steps back”*.

This finding also indicates that Joint Honour learners would be advised to spend the second half of the year abroad in the country where their weaker L2 is spoken. The rationale behind this is driven by the Skill Acquisition Theory (Anderson, 1982), which posits three stages of any skill acquisition. In an L2 which is more familiar to the learner, the time it takes for a skill to transfer from procedural knowledge to automated knowledge is lessened. Once a skill is automatised, the chances of making a mistake, unless fossilised, is also lessened (DeKeyser, 2007). From a linguistic perspective, learners have more chance of successfully transitioning from procedural knowledge to automated knowledge within one semester if they are already proficient (i.e., near automatised knowledge) before the sojourn experience. If successful, theoretically, learners would be less likely to make mistakes post-sojourn as evidenced by a higher global proficiency mark at post-test. However, it is questionable whether a semester is a sufficient length for a learner to successfully transition concerning the weaker L2. After all, this language requires more practice and input, and if learners fail to make the transition, they remain at the proceduralisation

stage, but due to no longer being in the L2 cannot advance due to lack of assumed input. By engaging with the weaker L2 second, a learner is potentially minimising the extent to which they are without practice in the weaker L2. This is because they are more likely to use the other L2 both in the 1st and 2nd half of the year abroad due to connections made in the 1st half. Moreover, this finding indicates that Joint Honour language learners should remain self-motivated and self-disciplined to seek out L2 materials not immediately available in the host community. This may be through maintaining connections in the first L2 community, watching and reading L2 material online using a VPN or academic institutions providing extra study materials for these learners.

When analysing Single Honour learners specifically, linguistic gain appeared to be more pronounced in the second half of the year. Such a finding goes against much of the literature available (e.g., Hessel, 2016; Rees & Klapper, 2007) which has demonstrated accelerated gains during the first three months abroad and then a plateauing effect. Given that a C-test typically measures general proficiency through the prism of accuracy, the accentuated gain in the latter half of the year may reflect language accuracy generally being acquired at a slower rate than that of fluency as identified in Study 1. This explanation is, however, questionable as both Hessel (2016) and Rees and Klapper (2007) both employed C-tests to measure general proficiency. A second possible explanation, as highlighted by the descriptive data, is that learners were not yet proficient enough to take advantage of the linguistic opportunities afforded to them in the first semester, but that in accordance with the Threshold Hypothesis (Cummins, 1976), once this threshold had been met (i.e., mid-sojourn), proficiency quickly improved. This theory has, however, typically been labelled at less proficient learners (DeKeyser, 2007), of which, given the relatively high scores found in the C-tests, language learners in Study 2 are unlikely to be considered of 'low proficiency'. As such, it would be sensible to hypothesise that learners were already passed this threshold before sojourning. In Rees and Klapper (2007), the average baseline score was 53.90 (out of 100), yet as aforementioned, the learners experienced rapid growth in the first three months and then a plateau. As such, a clear explanation for this finding is not immediately forthcoming.

This current study has also captured perceived linguistic competencies, and as such, has allowed for both objective and perceived competency measures to complement one another. Due to the lack of objectivity in self-reported measures, such instruments have long been criticised when used alone (Pellegrino, 1998). All four language skills were perceived to improve significantly,

and no skill was perceived to improve significantly more than another. Learners perceived the largest gain in listening, followed by speaking, reflecting the findings of Badstübner and Ecke (2009) who explored perceived change in 30 L1 English learners of German who partook in a summer exchange program, also finding learners to perceive the largest gains in listening and speaking. It may be hypothesised that these two skills improve the most during the year abroad because they are consistently actively used (Isabelli-García, Bown, Plews, & Dewey, 2018). Lastly, the study looked to ascertain whether personality could account for why some individuals experienced stronger change scores than others (RQ12). Personality has long been considered a *forgotten* variable in SLA (Dörnyei, 2005, Dörnyei & Ryan, 2015), but has nonetheless been considered an important factor in accounting for individual differences in linguistic gain (Dewaele, 2013). The study found that after controlling for baseline proficiency and Honours Programme, only that of baseline openness significantly predicted language gain. That is to say, the more open learners were at baseline, the higher their overall gain. This finding is in line with previous literature which has tended to base their conclusion on the relationship between openness and language gain driven by SLA theory, as opposed to directly testing for an association. For example, Ożańska-Ponikwia and Dewaele (2012) found openness was a significant predictor of frequency of use of L2 English, and therefore open learners were assumed to experience stronger L2 development. Similarly, individuals high in openness also tend to be interculturally competent (Byram, Nichols, & Steven, 2001; Deardoff, 2006). Individuals high in intercultural competence tend to find integration into new host communities easier, and with it find new channels for L2 interaction to take place (Taguchi, Xiao, Li, 2016). Consequently, these studies have argued that open learners are more successful language learners because they seek out opportunities to use the L2, which will, in turn, foster acquisition. Baker-Smemoe et al. (2014) is the only study, to the best of my knowledge, which has looked to investigate whether broad trait personality can account for individual differences found in language gain. The authors divided 102 students of differing L2s into *gainers* and *non-gainers*, finding broad baseline trait personality to be non-significantly different between the two groups as per a series of independent t-tests. Consequently, they concluded that personality was not a predictor of why some individuals gained more than others. However, these findings are of little robustness, and the regression analysis conducted in this current research has provided insights not possible from t-tests alone and have also allowed baseline proficiency and Honours programme to be controlled for, increasing the reliability of conclusions given.

6.2.4 Evaluating the experience as a whole

Through undertaking both a systematic review and a longitudinal, quasi-experimental study, this research has focused on the constructs of personality, well-being and language, to ascertain how sojourning may foster change in all three.

The thesis has shown sojourning appears to have multiplicative benefits on a range of measured constructs. Most pertinently, the thesis has indicated that the average individual returns home apparently significantly more agreeable, curious, and with higher global linguistic proficiency than if they were to remain at-home.

When assessing these changes over time, it is nonetheless, important to acknowledge the individual differences in the extent of personality change over time. While the year abroad can be a positive experience for some, represented by positive personal growth, for others, it can be a difficult and frustrating time. This study has drawn tentative conclusions regarding the factors influencing the extent to which sojourning fosters such changes. Through doing so, it provides an avenue for future intervention research and provides a point of discussion in pre-departure workshops. Moreover, the study has shown sojourning to appear to have little impact on psychological well-being, and over time, sojourners did not experience higher perceived well-being on average than those at-home.

From a linguistic perspective, Study 1 has provided an evaluative component, which was not possible in Study 2. The conclusions provided indicated that sojourning can facilitate improvement in oral fluency and general proficiency, but again given the methodological limitations in the field at large, these conclusions must be considered tentative.

In light of the research questions, it can be concluded that sojourning should be considered a valuable experience, an experience in which, in the right circumstances can foster personal and linguistic growth beyond that found in an at-home learning context. Institutions must, at the same time, however, appreciate that selling the year abroad as having a uniform effect may be misleading. While this study has made an important contribution into assessing the value of sojourning, perhaps upon reflection its most important contribution is in understanding how both stakeholders and the students themselves can maximise the potential of the sojourning experience so that it continues to meet its aims and objectives.

6.3 Implications

6.3.1 Implication for theoretical development

6.3.1.1 *The year abroad as a valid life event?*

Within the developmental literature, sojourning would be classed as a slow and continuous life event, one which changes behavioural responses over time. These transitions are less explored, but evidence in the past decade has suggested that personality can change within a short period (e.g., Bleidorn, 2012; Zimmermann & Neyer, 2013). The findings of the current study support the notions of Zimmermann and Neyer (2013) and Niehoff et al. (2017) in arguing that sojourning serves as a valid life event with the potential to influence personality change directly. While the aforementioned study found direct socialization effects for trait agreeableness, this study has found direct socialization effects for curiosity also. This provides evidence that sojourning can be classed as a life event which changes thoughts, feelings and behaviours not only in higher-order traits but also lower-order traits too.

6.3.1.2 *Principle of personality development*

As noted by Wolff, Schmidt, Borzikowsky, Möller and, Wagern (2020), the underlying mechanisms leading to personality change during a study abroad are debatable, and different explanations have been suggested. Using the Neo-Socioanalytic Model (Roberts & Wood, 2006) as a framework to discuss principles of personality change and stability, the findings of Study 2 are explained.

The Maturity Principle posits that with age, people become more socially dominant, agreeable, conscientious and emotionally stable (Roberts & Nickel, 2017). The most conclusive evidence in the current study that sojourning accelerates maturation is in the finding that sojourners became significantly more agreeable than those who remained at-home. Although less conclusive, due to the lack of control group, sojourners were also found to return home significantly more open and emotionally stable, again indicating that sojourners returned home with higher maturity than upon departure. There was little evidence to indicate that remaining at-home fostered maturation and supports the perspective that remaining in full-time education can inhibit maturity (Bleidorn & Schwaba, 2017).

Closely linked with that of the Maturity Principle, the Social Investment Principle posits as to *why* individuals mature with age (Roberts & Nickel, 2017): personality change is driven by the

requirement of individuals to commit to adult social roles (e.g., being responsible, minimising stress arousal) (Lodi-Smith & Roberts, 2007). In Study 2, the finding that those in employment during the year abroad appeared to become significantly more conscientious than those who studied abroad is evidence of this principle. As aforementioned above, full-time academic study will likely postpone entry into adult social roles (Arnett, 2000) and as such, inhibit changes in personality tendencies associated with maturity. Moreover, simply undertaking a sojourn may require learners to adopt adults' roles earlier than if they remained at-home. For example, in the descriptive data, learners spoke of becoming more independent, displaying greater desirability to become socially dominant, and having greater control in minimising neurotic arousal.

The Correspondive Principle posits that people enter into specific environments based on their specific trait make-up and that through living or experiencing these environments, these specific traits then develop. From a sojourning perspective, the few studies which have explored socialization effects (i.e., Zimmermann & Neyer, 2013; Niehoff et al., 2017) have explained these effects through the Correspondive Principle. For example, Zimmermann and Neyer (2013) found those who volunteered to study abroad not only reported higher initial levels of openness but also became significantly more open than their peers who remained at-home. From the perspective of Study 2, there was no explicit evidence that the Correspondive Principle accounts for trait trajectory development. After all, at baseline, sojourners significantly differed in trait neuroticism only, yet trajectories of change over time were not significantly different. One reason for this is that the sample predominantly consisted of students for whom the experience was compulsory, and as such, the presence of self-selection effects was minimised. Nevertheless, when exploring predictors of trait change, there was tentative evidence to suggest that the Correspondive Principle was present. In Study 2, for example, those who partook in extracurricular activities were found to become significantly more open and extraverted. Conversely, research has shown those who are more open and extraverted to be more likely to take part in extracurricular activities (e.g., Asendorpf & Wilpers, 1998).

6.3.1.3 Capturing variability and ascertaining its systematicity

As reviewed in Chapter 2, scholars have traditionally been interested in taking a nomothetic approach to broad trait personality, demonstrating a greater interest in the between-subject comparison. Advances in theory and methodology, however, mean the dynamic properties of personality are today of great interest.

This study has provided original insights into the properties of dynamic personality by exploring variability breakdown in a hierarchical framework of personality. Whereas studies typically focus on one level (e.g., broad traits), Study 2, has been the first to the best of my knowledge, to explore broad traits, narrow traits and states in a single framework. In doing so, it has provided important insights into how variability changes in response to the intensity of the repeated measure. This relationship has suggested that states are the most volatile of the three levels of personality, whereas broad and narrow traits are relatively more stable. It was an interesting insight that at each change of personality level within-person variability accounted for approximately 25% more of the total variability.

An important question in this research area has been whether within-person variability is at all meaningful or whether it should be ignored as error-variance. The findings have added to the growing body of literature that within-person variability in state personality (in Study 2, agreeableness) should not be considered as error variance but that this variability has meaning. This study has shown, that within-person variability is associated with changes in situational characteristics and demands, ultimately supporting the notion of situational contingencies (Fleeson, 2001; 2007; Minbashian et al., 2010; Sherman et al., 2015).

Lastly, Study 2 has aligned current theoretical perspectives regarding state personality within a new population. Past research has explored within-person variability in state personality in a range of contexts, such as a working context (e.g., Wood et al., 2019; Minbashian et al., 2010); an L2 classroom learning context (e.g., Zhang et al., 2018) and with undergraduate students partaking in at-home university study (e.g., Rauthmann et al., 2014; Fleeson, 2001). This is the first study, to the best of my knowledge, to explore state personality with a sojourning sample, which may display differing thoughts, feelings and behaviours from those in the aforementioned contexts. Given findings are similar to those of past research; this study has generalised findings to a sojourning sample.

6.3.2 Implications for policy and practitioners

The findings of the thesis had indicated several areas in which policy or practice could be adjusted in order to improve the sojourning experience.

6.3.2.1 *The mid-year dip*

When exploring well-being longitudinally, the average well-being score of sojourners appeared to dip in January and February. While small-scale, the focus group datasets have provided valuable insights into why this may be. Returning sojourners spoke of feeling isolated, lonely and demotivated during these winter months for the friends made in the first semester had returned their home university. Many spoke of having little motivation to go out and make new friends, while the motivation to explore new places had also lessened upon return. This experience may be specific to British sojourners, for sojourners of other countries tend to remain abroad for one semester only. From a home institution perspective, it is therefore advised, to offer extra support during this time. This may involve offering online support sessions, conducting more office hours or offering access to online resources such as the 'Big White Wall'. Moreover, home universities would be advised in pre-departure workshops to discuss possible difficulties post-Christmas. A possible task in these workshops could be to ask students to write a postcard detailing how they believe they will be feeling at this time and identify goals and objectives they wish to accomplish in the second half of the academic year. Personal goal setting and being attentive to these goals has been shown to foster well-being across a range of ages (e.g., Street et al., 2004; Grégoire, Bouffard, & Vezeau (2012).

6.3.2.2 *Reducing loneliness and promoting extracurricular participation*

Study 2 has explored how the sojourning experience differs among individuals and accounts for the factors which may influence the extent to which the experience is considered positive. Overall, two variables repeatedly appeared as predictors of the outcomes measured; that of loneliness and extracurricular participation.

The findings of Study 2 suggested an increased perception of loneliness inhibited positive personal growth and fostered poorer psychological well-being on average over time. As such, learners should be encouraged to build up a strong support network both before, and during the year abroad, as this may be vital in minimising loneliness. This may be achieved by implementing buddy schemes (e.g., Nieto & Nebot, 2020). While these buddy schemes are viewed typically through the prism of peer mentorship, participation in such schemes has been shown to foster support networks both at-home and abroad (University of Liverpool, 2020; Nieto & Nebot, 2020). Further research is required, however, to explore the explicit link between buddy schemes and loneliness.

Conversely, this current study has highlighted the importance of participation in extracurricular activities in both fostering personal development and psychological well-being. While pre-departure workshops advertise the benefits of extracurricular participation, reasoning has often not been empirically justified. Typically, past research has focused on the relationship between participation in extracurricular activities and intercultural competencies (e.g., Langley & Breese, 2005; Liu, 2016). This current study has been the first, to the best of my knowledge, to quantitatively examine the impact this participation can have on personality change. Learners must continue to be promoted to join societies and have easy access to societies in their local area. This could again be fostered through implementing a buddy scheme.

6.3.2.3 Challenging the notions of linguistic 'nativeness'

This current study urges learners and stakeholders to rethink the notion of 'nativeness' on a year abroad. Sojourning has long been thought of as the superior learning context (Hessel, 2017) and despite empirical rebuttal (e.g., Coleman & Chafer, 2010; Jackson, 2018), the year abroad is often sold as an immersive linguistic experience. Study 1 synthesised a broad range of SLA literature, concluding, only tentatively, that sojourning facilitates language learning at a greater rate than remaining at-home. Moreover, the extent of change was found to be highly compartmentalised with the strongest causal evidence of change being in the domain of general proficiency. Moreover, as noted in the descriptive dataset of Study 2, learners continually faced linguistic hurdles (e.g., English as a Lingua Franca, using the L2 solely in informal situations), which ultimately limited the extent to which they perceived linguistic change. Setting realistic linguistic goals is vital in ensuring learners do not return home disappointed and ultimately disillusioned with their linguistic progress (Badstübner & Ecke, 2009). This thesis had added to the growing body of literature which challenges the assumptions of study abroad and calls on practitioners to be more aligned to current empirical thinking when selling the year abroad experience.

6.4 Strengths and Limitations

6.4.1 Strengths

6.4.1.1 Sample make-up

Attempts were made to advertise and recruit participants from a range of disciplines in Study 2. Such a heterogeneous sample has answered the calls of previous research (e.g., Zimmermann & Neyer, 2013; Tracy-Ventura et al., 2016). As a result, Study 2 has been able to test for programme characteristics (e.g., compulsory/voluntary learner, location) when assessing factors

accounting for personality change and overall well-being. This focus had been missing in the available literature to date. Moreover, previous literature, particularly with regards to personality, has been undertaken on a German L1 speaking sample. As such, the findings presented have provided greater generalisability to the extent to which sojourning can foster outcome change. This is particularly important, given the specificities associated with the British sojourning experience (Alred & Byram, 2002). Furthermore, given the current political climate (i.e., Brexit) in the UK and its position in the ERASMUS programme, for example, research regarding British sojourners holds added validity and importance (see section 1.7) to policymakers.

6.4.1.2 Providing tentative causal conclusions regarding the value of sojourning

Typically, previous literature (e.g., McLeod & Wainwright, 2008; Tracy-Ventura et al., 2016) who have explored the value of sojourning on a series of outcomes, have done so by capturing data on an intervention group. Such a study design, however, results in an inability to causally attribute any observed change to the sojourning experience explicitly. A strength of the thesis as a whole is that each study has provided conclusions borne out of a quasi-experimental design, which given the moral and ethical implications employing an RCT, can be considered the most robust design in determining causality. Despite the limited sample size (see below), this thesis has, to the best of my knowledge, been the first to explore a series of outcomes (i.e., personality and well-being) within a British-domiciled sample which has looked to quantitatively attribute change in these outcomes to the study abroad. Given the specificities of the British sojourn experience (Alred & Byram, 2002), such research is required and in doing so can provide a platform from which policymakers and practitioners can empirically evaluate the value of the sojourning experience and inform future practice and preparation.

6.4.2 Limitations

Nonetheless, this thesis has methodological limitations which are important to evaluate when drawing conclusions.

6.4.2.1 Sample size

Each of the two studies has been adversely affected by small sample sizes. In Study 1, the sample sizes of the studies included in the systematic review ranged from 18 to 143, while in Study 2, the sample size of 180 is considerably smaller than other large-scale studies (e.g., Zimmermann & Neyer, 2013; Lüdkte et al., 2011). Having a small sample size is problematic for several reasons.

Firstly, in studies with small sample sizes, it can be difficult to ascertain whether the differences observed are due to the intervention (i.e., study abroad) or simply due to chance. Concerning the thesis as a whole, the sample sizes included were only large enough to detect both moderate and large-sized effects. Consequently, the threat of a type II error (i.e., concluding there is no effect when one exists) is higher than desirable due to some analyses being underpowered (Sullivan & Feinn, 2012).

Secondly, the limited sample size found in Study 2 imposed the number of variables which could be added to the final regression models. As noted by Field (2013), the sample size is important in regression if a reliable regression model is to be obtained. Given that different types of variables were tested (e.g., *person, environment*), it would have been preferable to add all variables into the model, so that the amount of variance explained by these could have been ascertained.

6.4.2.2 *Threat of potential biases*

6.4.2.2.1 *Sample bias*

In any research where participation is voluntary, questions arise regarding *who* chooses to partake in the study and how their participation may impact the final results. Those who consented to participation likely had an underlying motivation on why they chose to be studied. From a personality perspective, participants might be more curious or open to new ideas and perspectives. On the other hand, individuals may have underlying personality disorders or be more likely to experience depression and/or negative emotions. Moreover, from a linguistic proficiency perspective, individuals who participated may likely be more proficient and confident users of the L2 than those who did not participate, with this point holding relevance for studies included in the systematic review and Study 2.

Attempts were made to mitigate and control for the presence of sample bias. In Study 2, a financial incentive was provided in order to attract participants who may be extrinsically motivated to participate in the study. As noted by Hsieh and Kocielnik (2016), the type of study incentive influences the type of person attracted to the study, which may confound with the personality outcome in Study 2. Hsieh and Kocielnik (2016) found that studies which incorporated a lottery reward system (such as Study 2) attracted participants with stronger openness-to-change values, while a charity reward attracted those with higher values in self-transcendence. Moreover, in both Study 1 and Study 2, baseline measures were collected in order

to ascertain the extent to which the intervention and comparison groups differed at baseline on the outcome. The pre-test is an essential component in recognising the presence of sample bias (Christensen et al., 2014; Shadish et al., 2002). Lastly, in the regression analyses conducted in Study 2, baseline personality and linguistic measures were each controlled for when examining predictors of change.

Furthermore, while attempts were made to recruit learners from a range of universities, recruited participants are skewed towards those who studied at Russell Group universities (see Table 20). Therefore, findings hold little generalisability to students who studied at non-Russell Group universities. Over 15,000 students from a range of universities across Britain complete a study abroad each year. Consequently, the sample in Study 2 cannot be considered representative of a population as a whole.

6.4.2.2.2 Self-report bias

Response bias is a widely discussed phenomenon in behavioural research, which commonly uses self-report data (Rosenman, Tennekoon, & Hill, 2011). In the context of this study, the threat of self-report bias must be considered high as the topic under study was highly personal. Regarding the measures, an individual may wish to portray an 'ideal-self' (Rosenman et al., 2011; Jonge & Slaets, 2005) which is not corresponsive to one's real self (i.e., social-desirability bias). In order to meet the criterion of an individual's 'ideal-self', they may select higher in socially desirable characteristics, and consequently, skew the results of the questionnaire data. Individuals were asked to answer honestly, and without judgement, although one must be mindful that the characteristics selected may not be a true representation of one's tendencies. These cognitive biases can undoubtedly skew results and confound with self-report responses, and the reader must remain mindful of this when interpreting the findings.

6.4.2.2.3 Recall bias

Given that many of the questionnaires were completed after a set period, the reader must also be mindful of recall bias. Recall bias refers to the systematic error that can occur when individuals do not remember correctly or omit details about a particular event. The longer the length of time passed between this event occurring and the questionnaire being administered, the greater the chance of recall bias. The presence of recall bias is likely highest in the monthly questionnaires, whereby an individual's current thoughts, feelings and behaviours, consciously or subconsciously

may influence the responses provided. Again, this recall bias may skew results and provide an unreliable insight into the respondents' personality. Conversely, the state personality questionnaires demonstrated the strongest ecological validity for the data were *in the moment* and is, therefore, fairer representation of an individuals' thoughts, feelings and behaviours.

6.5 Future Directions

With the conclusion of this study, attention must now be focused on understanding how future studies can successfully build on this research.

6.5.1 Exploring the differences between compulsory and voluntary sojourners

Study 2 collated students for whom the experience was either compulsory or voluntary together. Nonetheless, as shown in the preliminary baseline analysis (see section 5.3.1), students who volunteered to study abroad were significantly different at baseline in broad trait neuroticism and all three measured narrow traits. Consequently, it would be interesting to identify whether compulsory students changed in trait personality at a differing rate to their volunteering peers. If substantial differences existed in the rate of change for traits such as neuroticism and curiosity, this would serve as evidence of the corresponive principle. Moreover, to offer fairer comparability with the findings of Zimmermann and Neyer (2013) and Niehoff et al. (2017), it would be of value to look explicitly at students who volunteered to partake in the study abroad. This would be an area of future research, given that the sample size collected in this study is too small to ensure that analyses are suitably powered.

6.5.2 Ensuring language studies hold methodological rigour

Through conducting the systematic review (Study 1), it became apparent how limited current study abroad literature is in relation to their methodological rigour. While it should be noted that scholars are limited in the study designs possible (e.g., given the moral and ethical challenges on randomly selecting students to partake in a year abroad; RCTs are highly unlikely to be employed), even where studies have contained a comparison group, reporting standards are relatively poor. For example, it may be unreported whether the two groups were equivalent at baseline, or in some instances, pre-testing may not be conducted at all. Such a criticism supports the finding of Tullock and Ortega (2017, p. 12), who found “in 44% of independent samples (11 out of 25), proficiency was either not reported or was uninterpretable based on the information that was provided.” This methodological rigour is required if we are to assess the linguistic

benefits of a sojourn suitably. As earlier noted in section 5.2.3, a potential alternative design to the traditional quasi-experimental design is a repeated measures design, where all participants receive all treatments (study abroad and at-home). While these studies were captured in the mapping table of Study 1, they were not brought forward to the in-depth review due to a desire to include a control group to compare results with a counterfactual, stay-at-home comparison group.

6.5.3 Re-entry and long-term personality change

Whereas aspects of culture shock and entry into a host community have been documented (e.g., Bathke & Kim, 2016), to date, little research has investigated re-entry into the home community (Wielkiewicz & Turkowski, 2010; Pelletier, 2019). Zimmermann and Neyer (2013) proposed the process of re-entry as serving as a valid life event within the prism of the personality literature, but further research is required to examine this. Questions also remain regarding whether the observed personality changes during the year abroad are retained over a longer period. Richter, Zimmermann, Neyer and Kandler (2020) undertook a follow-up study from Zimmermann & Neyer (2013). Capturing personality change in 441 participants who partook in the original study over five years, the findings indicated that the observed sojourn effects for openness and neuroticism remained over time. Agreeableness scores were found to decrease however, after return home, providing further tentative evidence that a sojourn experience has a particular impact on trait agreeableness. Further evidence is therefore required to ascertain external validity with these findings, and given the specificities of the British sojourn experience, questions remain over the long-term effects of a longer stay abroad.

6.6 Conclusions

Each study aimed to evaluate the value of sojourning on an individual's personal and linguistic development compared to remaining at-home. While the samples presented in each study must be considered convenient and therefore, not indicative of the population at large, the findings have provided both a descriptive and explanatory account of observed outcome changes.

The findings suggest that sojourning can be a valuable experience for many who undertake it. There appeared a strong additive effect on agreeableness and curiosity, indicating that these are the two personality constructs which benefit the most from the experience. Nonetheless, the year abroad must be considered as a highly individualised experience, and some will see greater

benefit from it than others. As such, the study has also captured predictive factors in the extent of this development, identifying loneliness to significantly impact personal development in both those who go abroad and stay at-home. From a linguistic perspective, sojourning appeared to facilitate language development, but only when living in the host country of the learnt target language. This development in global proficiency, as seen in Study 2, supports the findings of the systematic review, which found global proficiency also developed while abroad in those studies considered most methodologically rigorous. Concerning predictors of language change, the study found baseline openness and degree programme to impact the extent of change made. The study has also provided ideas for future studies and how this study may be used to move policy forward and improve the learning experience of 3rd year learners regardless of learning context. Most importantly, it is hoped that these results can better help realise student expectations of the year abroad experience, in order to minimise learner disappointment and disillusionment with the experience upon return home.

References

- Abdellaoui, A., Chen, H. Y., Willemsen, G., Ehli, E. A., Davies, G. E., Verweij, K. J., ... & Cacioppo, J. T. (2019). Associations between loneliness and personality are mostly driven by a genetic association with neuroticism. *Journal of Personality*, 87(2), 386-397.
- Alharbi, E. S., & Smith, A. P. (2018). Review of the literature on stress and well-being of international students in English-speaking countries. *International Education Studies*, 11(6), 22-44.
- Allen, H. (2002). *Does study abroad make a difference? an investigation of linguistic and motivational outcomes*. (unpublished doctoral dissertation). Emory University: United States.
- Allen, H. (2010). What shapes short-term study abroad experiences? a comparative case study of students' motives and goals. *Journal of Studies in International Education*, 14(5), 452-470.
- Allen, M., Magee, C., Vella, S., & Laborde, S. (2017). Bidirectional associations between personality and physical activity in adulthood. *Health Psychology*, 36, 332-336.
- Allison, P.D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, 20, 93-114.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. London: Constable.
- Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt, Reinhart & Winston.
- Alred, G., & Byram, M. (2002). Becoming an intercultural mediator: A longitudinal study of residence abroad. *Journal of Multilingual and Multicultural Development*, 23(5), 339-352.
- Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369-406.
- Anderson, P. H., Lawton, L., Rexeisen, R. J., & Hubbard, A. C. (2006). Short-term study abroad and intercultural sensitivity: A pilot study. *International Journal of Intercultural Relations*, 30(4), 457-469.
- Andrews, B., & Wilding, J. M. (2004). The relation of depression and anxiety to life-stress and achievement in students. *British Journal of Psychology*, 95(4), 509-521.
- Andrews, G., Page, A. C., & Neilson, M. (1993). Sending your teenagers away: Controlled stress decreases neurotic vulnerability. *Archives of General Psychiatry*, 50(7), 585-589.
- Argyle, M., & Lu, L. (1990). The happiness of extraverts. *Personality and Individual Differences*, 11, 1011-1017.
- Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, 55, 469-480.
- Arvidsson, K., Eyckmans, J., Rosiers, A., & Lundell, F. F. (2018). Self-perceived linguistic progress, target language use and personality development during study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 3(1), 144-166.
- Asendorpf, J. B. (2009). Personality: Traits and situations. In J. P. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (pp. 43-53). Cambridge: Cambridge University Press.
- Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, 74(6), 1531-1544.
- Ashton, M.C., Lee, K., Penugini, M., Szarota, P., Devries, R. E., Di Bias, L., Boates, K., & De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, 86, 356-366.
- Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology*, 68(6), 627-636.
- Avello, P. (2014). *L2 phonological development in speech production during study abroad* (unpublished doctoral dissertation). Universitat Pompeu Fabra: Barcelona.

- Avello, P., & R. Lara. (2014). Phonological development in speech production and study abroad programmes varying in length. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 137–167). Amsterdam: John Benjamins.
- Avello, P., Mora, J., & Pérez-Vidal, C. (2012). Perception of FA by non-native listeners in a study abroad context. *Research in Language*, 10(1), 63-78.
- Backteman, G., & Magnusson, D. (1981). Longitudinal stability of personality characteristics. *Journal of Personality*, 49(2), 148-160.
- Bacon, S. M. (2002). Learning the rules: Language development and cultural adjustment during study abroad. *Foreign Language Annals*, 35(6), 637-646.
- Badstübner, T., & Ecke, P. (2009). Student expectations, motivations, target language use, and perceived learning progress in a summer study abroad program in Germany. *Teaching German*, 42(1), 41-49.
- Bakalis, S., & Joiner, T. (2004). Participation in tertiary study abroad programmes: The role of personality. *The International Journal of Educational Management*, 18(5), 286-291.
- 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.
- Bakker, W., Van Oudenhoven, J. P., & Van Der Zee, K. I. (2004). Attachment styles, personality, and Dutch emigrants' intercultural adjustment. *European Journal of Personality*, 18(5), 387-404.
- Bardi, A., & Ryff, C. D. (2007). Interactive effects of traits on adjustment to a life transition. *Journal of Personality*, 75, 955-983.
- Barquin, E. L. (2012). *Writing development in a study abroad context* (unpublished doctoral dissertation). Universitat Pompeu Fabra: Barcelona.
- Barron, A. (2019). Using corpus-linguistic methods to track longitudinal development: Routine apologies in the study abroad context. *Journal of Pragmatics*, 146, 87-105.
- Bateman, B. (2002). Promoting openness toward culture learning: Ethnographic interviews for students of Spanish. *Modern Language Journal*, 86, 318–331.
- Bathke, A., & Kim, R. (2016). Keep calm and go abroad: The effect of learning abroad on student mental health. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 27, 1-16.
- Beattie, J., Valls-Ferrer, M., & Pérez-Vidal, C. (2014). Listening performance and onset level in formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 87–110). Amsterdam: John Benjamins.
- Beaven, A., & Spencer-Oatey, H. (2016). Cultural adaptation in different facets of life and the impact of language: A case study of personal adjustment patterns during study abroad. *Language and Intercultural Communication*, 16(3), 349-367.
- Beckmann, N., Wood, R.E., & Minbashian, A (2010). It depends how you look at it: On the relationship between neuroticism and conscientiousness at the within-and the between-person levels of analysis. *Journal of Research in Personality*, 44(5), 593-601.
- Beckmann, N., & Wood, R. E. (2017). Dynamic personality science. Integrating between-person stability and within-person change. *Frontiers in Psychology*, 8, 1486.
- Beckmann, N. & Wood, R.E. (2020). Dynamic personality at work. In Y. Griep, S.D. Hansen, T. Vantilborgh, & J. Hofmans (Eds.), *Handbook on The Temporal Dynamics of Organizational Behavior* (pp. 197 – 220). Cheltenham: Edward Elgar Publishing.
- Belnap, R. C., Bown, J., Dewey, D. P., Belnap, L. P., & Steffen, P. R. (2016). Project perseverance: Helping students become self-regulating learners. In P. D. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive Psychology in SLA* (pp. 282-304). Bristol: Multilingual Matters.
- Beltz, A. M., Wright, A. G., Sprague, B. N., & Molenaar, P. C. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment*, 23(4), 447-458.
- BERA. (2018). *Ethical guidelines for educational research (3rd edition)*. Retrieved March 5, 2018, from <https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2011>

- Buecker, S., Maes, M., Denissen, J. J., & Luhmann, M. (2020). Loneliness and the big five personality traits: A meta-analysis. *European Journal of Personality, 34*(1), 8-28.
- Bewick, B.M., J. Gill, B. Mulhern, M. Barkham, And A.J. Hill. (2008). Using electronic surveying to assess psychological distress within the UK university student population: A multisite pilot investigation. *E-Journal of Applied Psychology, 4*, 1-5.
- Bewick, B., Koutsopoulou, G., Miles, J., Slaa, E., & Barkham, M. (2010). Changes in undergraduate students' psychological well-being as they progress through university. *Studies in Higher Education, 35*(6), 633-645.
- Blackie, L. E., Roepke, A. M., Forgeard, M. J., Jayawickreme, E., & Fleeson, W. (2014). The promise of changing personality states to promote well-being. In S. Schueller & A. Parks (Eds.), *The Wiley Blackwell Handbook of Positive Psychological Interventions* (pp. 462-474). Chichester, West Sussex, United Kingdom.
- Bleasdale, L. & Humphreys, S. (2018). *undergraduate resilience research report*. Retrieved March 1, 2019, from https://teachingexcellence.leeds.ac.uk/wp-content/uploads/sites/89/2018/01/litebleasdalehumphreys_fullreport_online.pdf
- Bleidorn, W. (2012). Hitting the road to adulthood: Short-term personality development during a major life transition. *Personality and Social Psychology Bulletin, 38*(12), 1594-1608.
- Bleidorn, W. (2015). What accounts for personality maturation in early adulthood? *Current Directions in Psychological Science, 24*(3), 245-252.
- Bleidorn, W., Klimstra, T. A., Denissen, J. J. A., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2013). Personality maturation around the world: A cross-cultural examination of social-investment theory. *Psychological Science, 24*(12), 2530–2540.
- Bleidorn, W., Schwaba, T. (2017). Personality development in emerging adulthood. In Specht, J. (Eds.), *Personality Development Across the Lifespan*, (pp. 39-52). London: Elsevier.
- Bleidorn, W., Hopwood, C., & Lucas, R. (2018). Life events and personality trait change. *Journal of Personality, 86*(1), 83-96.
- Bliese, P. (2006). *Multilevel modeling in R (2.2) A brief introduction to R, the multilevel package and the nlme package*. Retrieved March 15, 2019, From https://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf
- Bolger, N., & Laurenceau, J. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford Press.
- Borràs, J., & Llanes, À. (2019). Re-examining the impact of study abroad on L2 development: A critical overview. *The Language Learning Journal, 1*-14.
- Bornstein, M., Jager, J., & Putnick, D. (2013). Sampling in developmental science: situations, shortcomings, solutions, and standards. *Developmental Review, 33*(4), 357-370.
- Bouchard, T. J., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology, 54*(1), 4-45.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in Psychology. *Qualitative Research in Psychology, 3*(2), 77-101.
- British Council (2005). *Breaking the barriers: 100 years of the language assistants programme 1905-2005*. London: British Council.
- British Council. (2017). *Broadening Horizons*. Retrieved August 10, 2018, from <https://www.britishcouncil.org/education/ihe/knowledge-centre/student-mobility/broadening-horizons-2017-addressing-needs-new>
- Brown, L., & Holloway, I. (2008). The initial stage of the international sojourn: Excitement or culture shock? *British Journal of Guidance & Counselling, 36*(1), 33-49.
- Brown, N., Neel, R., & Sherman, R. (2015). Measuring the evolutionarily important goals of situations: Situational affordances for adaptive problems. *Evolutionary Psychology, 13*(3), 1-15.
- Bryman, A. (2012). *Social research methods* (4th ed.). Oxford: Oxford University Press.
- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

- Byram, M. S. (2008). Researching residence and study abroad. In S. Ehrenreich, G. Woodman & M. Perrefort (Eds.), *Auslandsaufenthalte in Schule und Studium. Bestandsaufnahme aus Forschung und Praxis* (pp. 19 – 28). Münster: Waxman Verlag.
- Cacioppo, J., Hawkley, L., Ernst, J., Burleson, M., Berntson, G., Nouriani, B., & Spiegel, D. (2006). Loneliness within a nomological net: an evolutionary perspective. *Journal of Research in Personality, 40*(6), 1054-1085.
- Campbell, D., & Stanley, J. (1966). *Experimental and quasi-experimental designs for research*. London: Houghton Mifflin Company.
- Campbell, M., McKenzie, J., Sowden, A., Katikireddi, S., Brennan, S., Ellis, S., . . . Thomson, H. (2020). Synthesis without meta-analysis (SWIM) in systematic reviews: Reporting guideline. *BMJ, 368*, 1-6.
- Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals, 1*, 131–151.
- Carpenter, J., & Plewis, I. (2011). Analysing longitudinal studies with non-response: Issues and statistical methods. In W. M. Vogt, (Ed.), *The SAGE handbook of innovation in social research methods* (pp. 498-523). New York: Sage.
- Carsello, C., & Creaser, J. (1976). How college students change during study abroad. *College Student Journal, 10*, 276-278.
- Carter, N., Guan, L., Maples, J., Williamson, R., & Miller, J. (2016). The downsides of extreme conscientiousness for psychological well-being: The role of obsessive-compulsive tendencies. *Journal of Personality, 84*(4), 510-522.
- Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Reviews. Psychology, 56*, 453-484.
- Centre for Reviews and Dissemination. (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. Accessed on January 5, 2020, from https://www.york.ac.uk/media/crd/systematic_reviews.pdf
- Chalmers, I., Hedges, L., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & The Health Professions, 25*(1), 12-37.
- Chapell, M., Blanding, Z., Silverstein, M., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology, 97*(2), 268-274.
- Chan, J. C., & Sy, P. Y. (2016). The relationships among personality, intercultural communication, and cultural self-efficacy in nursing students. *The Journal of Nursing Research, 24*(4).
- Charney, D. (2003). The psychobiology of resilience and vulnerability to anxiety disorders: Implications for prevention and treatment. *Dialogues in Clinical Neuroscience, 5*(3), 207-221.
- Chavajay, P., & Skowronek, J. (2008). Aspects of acculturation stress among international students attending a university in the USA. *Psychological Reports, 103*(3), 827–835.
- Chen, F. F., Jing, Y., Hayes, A., & Lee, J. M. (2013). Two concepts or two approaches? A bifactor analysis of psychological and subjective well-being. *Journal of Happiness Studies, 14*(3), 1033-1068.
- Cheng, Y. S. (2002). Factors associated with foreign language writing anxiety. *Foreign Language Annals, 35*(6), 647-656.
- Chieffo, L., & Griffiths, L. (2004). Large-scale assessment of student attitudes after a short-term study abroad program. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10*, 165-177.
- Ching, C. M., Church, A. T., Katigbak, M. S., Reyes, J. A. S., Tanaka-Matsumi, J., Takaoka, S., Zhang, H., Shen, J., Arias, R. M., Rincon, B. C., & Ortiz, F. A. (2014). The manifestation of traits in everyday behavior and affect: a five-culture study. *Journal of Research in Personality, 48*, 1–16.

- Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014). Understanding quantified-selfers' practices in collecting and exploring personal data. *Proceedings of The SIGCHI Conference on Human Factors in Computing Systems* (pp. 1143–1152). New York, NY: Association for Computing Machinery.
- Christensen, L., Johnson, B., & Turner, L. (2014). *Research methods, design, and analysis* (12th ed.). Harlow: Pearson.
- Churchill, E., & Dufon, M. (2006). *Language learners in study abroad contexts*. Blue Ridge Summit: Multilingual Matters.
- Clarke, M., & Chalmers, I. (2018). Reflections on the history of systematic reviews. *BMJ Evidence-Based Medicine*, 23(4), 121-122.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education* (7th ed.). Hoboken: Taylor And Francis.
- Cooke, R., Bewick, B.M., Barkham, M., Bradley, M., & Audin, K. (2006). Measuring, monitoring and managing the psychological well-being of first year university students. *British Journal of Guidance and Counselling*, 34, 505–17.
- Coleman, J.A. (1996). *Studying languages: A survey of British and European students*. London: Centre for Information on Language Teaching and Research.
- Coleman, J. A. (1998). Language learning and study abroad: The European perspective. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 4(2), 167–203.
- Coleman, J. (2013). Researching whole people and whole lives. In C. Klinger (Ed.), *Social and Cultural Aspects of Language Learning in Study Abroad* (pp. 17 – 45). Amsterdam: John Benjamins.
- Coleman, J. A. (2015). Social circles during residence abroad: What students do, and who with. *Social Interaction, Identity and Language Learning During Residence Abroad*, 4, 33-52.
- Coleman, J. A. & T. Chafer (2010). Study abroad and the internet: Physical and virtual context in an era of expanding telecommunications. *Frontiers: The Interdisciplinary Journal of Study Abroad* 19, 151–167.
- Collentine, J. (2009). Study abroad research: Findings, implications and future directions. In C. Doughty & M. Long (Eds.), *Handbook of Language Teaching* (pp. 218–233). Malden, MA: Blackwell.
- Conner, T. S., & Lehman, B. (2012). Getting started: Launching a study in daily life. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 89-107). New York: Guilford Press.
- Cooper, H. (2017). *Research synthesis and meta-analysis: a step-by-step approach* (5th ed.). Los Angeles: Sage.
- Cooper, H., Hedges, L., & Valentine, J. (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.
- Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory*. Odessa, FL: Psychological Assessment Resources.
- Council of Ministers (1987). Council decision of 15 June 1987 adopting the European community action scheme for the mobility of university students (ERASMUS). *Official Journal of The European Communities*, 166, 20–24.
- Cubillos, J., & Ilvento, T. (2012). The impact of study abroad on students' self-efficacy perceptions. *Foreign Language Annals*, 45(4), 494-511.
- Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*. 9, 1–43.
- Dalecki, M., & Willits, F. K. (1991). Examining change using regression analysis: Three approaches compared. *Sociological Spectrum*, 11(2), 127-145.

- Darling, N., Caldwell, L.L., & Smith, R. (2005). Participation in school-based extracurricular activities and adolescent adjustment. *Journal of Leisure Research*, 37, 51–76.
- Davidson, D. E. (2010). Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals*, 43(1), 6-26.
- Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, 10(3), 241-266.
- Deardorff, D. (2009). Preface. In D. Deardorff (Ed.). *The SAGE handbook of intercultural competence* (pp. xi – xiv). Los Angeles, CA: Sage.
- De Boer, M., Waterlander, W., Kuijper, L., Steenhuis, I., & Twisk, J. (2015). Testing for baseline differences in randomized controlled trials: An unhealthy research behavior that is hard to eradicate. *The International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 4-12.
- De Jonge, P., & Slaets, J. P. J. (2005). Response sets in self-report data and their associations with personality traits. *The European Journal of Psychiatry*, 19(4), 209–214.
- DeKeyser, R. (2007). Study abroad as foreign language practice. In DeKeyser, R. (Ed.). (2007). *Practice in a second language: perspectives from applied linguistics and cognitive psychology* (pp. 208 – 226). Cambridge: Cambridge University Press.
- DeKeyser, R. M. (2014). Research on language development during study abroad. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 313-326). Amsterdam: John Benjamins.
- Deneve, K.M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, 124, 197-229.
- Denovan, A., & Macaskill, A. (2017). Stress and subjective well-being among first year UK undergraduate students. *Journal of Happiness Studies*, 18(2), 505-525.
- Dewaele, J. M. (2012). Personality: Personality traits as independent and dependent variables. In S. Mercer, S. Ryan, & M. Williams (Eds.), *Psychology for Language Learning*. Palgrave Macmillan, London.
- Dewaele, J. M. (2013). The link between foreign language classroom anxiety and psychoticism, extraversion, and neuroticism among adult bi-and multilinguals. *The Modern Language Journal*, 97(3), 670-684.
- Dewaele, J. M., & Furnham, A. (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences*, 28(2), 355-365.
- Dewaele, J. M., & Thirtle, H. (2009). Why do some young learners drop foreign languages? A focus on learner-internal variables. *International Journal of Bilingual Education and Bilingualism*, 12(6), 635-649.
- Dewaele, J.M., Comanaru, R. & Faraco, M. (2015). The affective benefits of a pre-session course at the start of study abroad. In Mitchell, R., McManus, K., & Tracy-Ventura, N. (Eds.), *Social Interaction, Identity and Language Learning During Residence Abroad* (pp. 95-114). EUROSLA Monographs Series 4.
- 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.
- Dewey, D. P., Bown, J., & Eggett, D. (2012). Japanese language proficiency, social networking, and language use during study abroad: Learners' perspectives. *Canadian Modern Language Review*, 68, 111–137.
- Dewey, D. P., 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 programmes: An exploratory analysis of possible predictors. *Language Learning, 64*(1), 36-71.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology, 93*(5), 880-896.
- Diao, W., Freed, B., & Smith, L. (2011). Confirmed beliefs or false assumptions? A study of home stay experiences in the French study abroad context. *Frontiers: The Interdisciplinary Journal of Study Abroad, 21*, 109-142.
- Donatelli, L. (2010). The impact of technology on study abroad. In W. Hoffa & S. DePaul (Eds.), *A History of U.S. Study Abroad: 1965–Present*. Carlisle, PA: *Frontiers: The Interdisciplinary Journal of Study Abroad*, Special Issue, 295–320.
- Donnellan, M. B., Conger, R. D., & Burzette, R. G. (2007). Personality development from late adolescence to young adulthood: Differential stability, normative maturity, and evidence for the maturity-stability hypothesis. *Journal of Personality, 75*, 237-264.
- Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging, 23*(3), 558–566.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum Associates Publishers.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge.
- Dreary, J. I. (2009). The trait approach to personality. In J. Corr & G. Matthews (Eds.), *The Cambridge Handbook of Personality Psychology* (pp. 89-109). Cambridge: Cambridge University Press.
- Driskell, J. E., Goodwin, G. F., Salas, E., & O'Shea, P. G. (2006). What makes a good team player? Personality and team effectiveness. *Group Dynamics: Theory, Research, and Practice, 10*, 249– 271.
- Earnest, D. R., Rosenbusch, K., Wallace-Williams, D., & Keim, A. C. (2016). Study abroad in psychology: Increasing cultural competencies through experiential learning. *Teaching of Psychology, 43*(1), 75-79.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-Tests. *Language Testing, 23*(3), 290-325.
- Edmonds, A., & Gudmestad, A. (2018). Gender marking in written L2 French: Before, during, and after residence abroad. *Study Abroad Research in Second Language Acquisition and International Education, 3*, 59-82.
- Edwards, J., & Templeton, A. (2005). The structure of perceived qualities of situations. *European Journal of Social Psychology, 35*(6), 705-723.
- Ehrman, M. E. (1996). *Understanding second language learning difficulties*. Thousand Oaks, CA: Sage.
- Ehrman, M. E. (2008). Personality and good language learners. *Lessons from Good Language Learners*, 61-72.
- Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal, 89*(2), 206-220.
- Ellis, R. (1994). *The study of Second Language Acquisition*. Oxford: Oxford University Press.
- Elpidorou, A. (2017). The good of boredom. *Philosophical Psychology, 31*(3), 323-351.
- ERASMUS+. (2019). *The transition period*. Retrieved March 15, 2020, from <https://www.ERASMUSplus.org.uk/the-transition-period>
- Engle, L., Engle, J. (2003). Study abroad levels: Towards a classification of program types. *Frontiers: The Interdisciplinary Journal of Study Abroad, 9*, 1-20.
- Engel, C. (2010). the impact of ERASMUS mobility on the professional career: Empirical results of international studies on temporary student and teaching staff mobility. *Belgeo. Revue Belge De Géographie, 4*, 351-363.

- European Commission (2015). *on the way to ERASMUS+. a statistical overview of the ERASMUS programme in 2012-13*. Luxembourg: Publications Office of The European Union.
- European Commission. (2017). *ERASMUS+ annual report 2016 (statistical annex)*. Retrieved July 1, 2018, from https://ec.europa.eu/programmes/erasmus-plus/resources/documents/erasmus-annual-report-2016_en
- European Commission. (2019). *ERASMUS+ annual report 2017 (statistical annex)*. Retrieved August 12, 2019, from <https://publications.europa.eu/en/publication-detail/-/publication/519aa03d-1f0b-11e9-8d04-01aa75ed71a1/language-en/format-pdf/source-search>
- Evans, C., Connell, J., Audin, K., Sinclair, A., & Barkham, M. (2005). Rationale and development of a general population well-being measure: Psychometric status of the GP-Core in a student sample. *British Journal of Guidance & Counselling*, *33*(2), 153-173.
- Eysenck, H.J. (1992). Four ways five factors are *not* basic. *Personality and Individual Differences*, *13*, 667-673.
- Fayard, J. V., Roberts, B. W., Robins, R. W., & Watson, D. (2012). Uncovering the affective core of conscientiousness: The role of self-conscious emotions. *Journal of Personality*, *80*(1), 1-32.
- Félix-Brasdefer, J. C. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language Learning*, *54*(4), 587-653.
- Félix-Brasdefer, J. C., & Hasler-Barker, M. (2015). Complimenting in Spanish in a short-term study abroad context. *System*, *48*, 75-85.
- Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Thousand Oaks: SAGE Publications.
- Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior*, *33*(7), 865-877.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distribution of states. *Journal of Personality and Social Psychology*, *80*(6), 1011-1027.
- Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, *75*(4), 825-862.
- Fleeson, W., & Nofle, E. (2008). The end of the person–situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, *2*(4), 1667-1684.
- Fleeson, W., & Gallagher, P. (2009). The implications of big five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*(6), 1097–1114.
- Fong, J. (2020). An evaluation of an education abroad program on the intercultural learning and cross-cultural adaptability skills of university undergraduates. *Higher Education Evaluation and Development*. 55-68.
- Forsey, M., Broomhall, S., & Davis, J. (2012). Broadening the mind? Australian student reflections on the experience of overseas study. *Journal of Studies in International Education*, *16*(2), 128-139.
- Freed, B.F. (1995). What makes us think that students who study abroad become fluent? In B.F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam: John Benjamins.
- 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 programmes. *Studies in Second Language Acquisition*, *26*(02), 275-301.
- Friedman, H.S., Kern, M.L., & Reynolds, C.A. (2010). Personality and health, subjective well-being, and longevity. *Journal of Personality*, *78*, 179–216.
- Funder, D. C. (2001). Personality. *Annual Review of Psychology*, *52*, 197-221.
- Funder, D. C. (2006). Towards a resolution of the personality triad: persons, situations, and behaviors. *Journal of Research in Personality*, *40*(1), 21-34.

- Funder, D. C. (2008). Persons, situations, and person-situation interactions. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (pp. 568–580). New York: The Guilford Press.
- Furnham, A., & Bochner, S. (1986). *Culture shock: Psychological reactions to unfamiliar environments*. New York: Methuen.
- Furnham, A., Zhang, J., & Chamorro-Premuzic, T. (2006). The relationship between psychometric and self-estimated intelligence, creativity, personality and academic achievement. *Imagination, Cognition and Personality, 25*(2), 119–145.
- Furukawa, T., & Shibayama, T. (1993). Predicting maladjustment of exchange students in different cultures: A prospective study. *Social Psychiatry and Psychiatric Epidemiology, 28*(3), 142–146.
- Gass, S. & Selinker, L. (2008). *Second language acquisition: An introductory course*. Hillsdale: Lawrence Erlbaum Associates.
- Geeraert, N., Li, R., Ward, C., Gelfand, M., & Demes, K. (2019). A tight spot: How personality moderates the impact of social norms on sojourner adaptation. *Psychological Science, 30*(3), 333–342.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Gerpott, F., Balliet, D., Columbus, S., Molho, C., De Vries, R., & Cooper, M. L. (2018). How do people think about interdependence? A multidimensional model of subjective outcome interdependence. *Journal of Personality and Social Psychology, 115*(4), 716–742.
- Gieser, J. D. (2015). A sociocultural investigation of identity: How students navigate the study abroad experience. *Journal of College Student Development, 56*(6), 637–643.
- Golle, J., Rose, N., Göllner, R., Spengler, M., Stoll, G., Hübner, N., ... Nagengast, B. (2019). School or work? The choice may change your personality. *Psychological Science, 30*(1), 32–42.
- Gomes Da Costa, B., Smith, T. M. F., & Whitely, D. (1975). *German language attainment: A sample survey of universities and colleges in the uk*. Heidelberg: Julius Gross Verlag.
- Gorard, S. (2007). The dubious benefits of multi-level modelling. *International Journal of Research & Method in Education, 30*(2), 221–236.
- Gorard, S. (2013). *Research design: Creating robust approaches for the social sciences*. London, SAGE.
- Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd Ed.). Los Angeles: SAGE.
- Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*(3), 499–510.
- Grégoire, S., Bouffard, T., & Vezeau, C. (2012). Personal goal setting as a mediator of the relationship between mindfulness and well-being. *International Journal of Well-Being, 2*(3), 236–250.
- Greischel, H., Noack, P., & Neyer, F. J. (2016). Sailing uncharted waters: adolescent personality development and social relationship experiences during a year abroad. *Journal of Youth and Adolescence, 45*(11), 2307–2320.
- Grey, S. (2018). Quantitative approaches for study abroad research. In C. Sanz & A. Morales-Front (Eds.), *The Routledge Handbook of Study Abroad Research and Practice* (pp. 48–57). New York: Routledge.
- Grotjahn, R. (2002) Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis. In R. Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und Praktische Anwendungen, Band 4* (pp. 211–225). Bochum: AKS Verlag.
- Gu, Q., & Maley, A. (2008). Changing places: A study of Chinese students in the UK. *Language and Intercultural Communication, 8*(4), 224–245.
- Gu, Q., Schweisfurth, M., & Day, C. (2010). Learning and growing in a ‘foreign’ context: intercultural experiences of international students. *Compare, 40*(1), 7–23.

- Guilmette, M., Mulvihill, K., Villemaire-Krajden, R., & Barker, E. (2019). Past and present participation in extracurricular activities is associated with adaptive self-regulation of goals, academic success, and emotional well-being among university students. *Learning and Individual Differences*, 73, 8-15.
- Haas, B. W. (2018). The impact of study abroad on improved cultural awareness: A quantitative review. *Intercultural Education*, 29(5-6), 571-588.
- Haddadi, P., & Besharat, M. A. (2010). Resilience, vulnerability and mental health. *Procedia-Social and Behavioral Sciences*, 5, 639-642.
- Hadis, B. F. (2005). Gauging the impact of study abroad: how to overcome the limitations of a single-cell design. *Assessment & Evaluation in Higher Education*, 30(1), 3-19.
- Hanna, C., Tefertiller, S., & Cota, S. (2015). pepperdine one semester abroad: group satisfaction & emotional well-being. *Pepperdine Journal of Communication Research*, 3(1), 4-25.
- Harari, G. M., Müller, S. R., Mishra, V., Wang, R., Campbell, A. T., Rentfrow, P. J., & Gosling, S. D. (2017). An evaluation of students' interest in and compliance with self-tracking methods: Recommendations for incentives based on three smartphone sensing studies. *Social Psychological and Personality Science*, 8(5), 479-492.
- Härkönen, A., & Dervin, F. (2016). Study abroad beyond the usual 'imagineering'? the benefits of a pedagogy of imaginaries. *East Asia*, 33(1), 41-58.
- Harms, P. D. (2019). Corresponsive principle of personality development. In V. Zeigler-Hill, & T. Shackelford (Eds.), *Encyclopaedia of Personality and Individual Differences*. Cham: Springer.
- Harrell, F. E. (2001). *Regression modelling strategies*. New York: Springer.
- Hastings, A. J. (2002). In defense of C-Testing. In R. Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und Praktische Anwendungen* (pp. 11-25). Bochum: AKS-Verlag.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.
- Heller, D., Komar, J., & Lee, W. B. (2007). The dynamics of personality states, goals, and well-being. *Personality and Social Psychology Bulletin*, 33(6), 898-910.
- Hensley, B., Martin, P., Margrett, J. A., MacDonald, M., Siegler, I. C., & Poon, L. W. (2012). Life events and personality predicting loneliness among centenarians: Findings from the Georgia centenarian study. *Journal of Psychology*, 146, 173-188.
- 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.
- 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* (Unpublished Doctoral Dissertation). Oxford University: United Kingdom.
- Hessel, G. (2017). A new take on individual differences in L2 proficiency gain during study abroad. *System*, 66, 39-55.
- Hessel, G., & Vanderplank, R. (2018). What difference does it make? Examining English proficiency gain as an outcome of participation in ERASMUS study abroad programmes in the UK. *Study Abroad Research in Second Language Acquisition and International Education*, 3(2), 191-219.
- Higgins, J. P. T., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. Sussex: John Wiley & Sons.
- Hogan, R. (2009). Much ado about nothing: The person-situation debate. *Journal of Research in Personality*, 43(2), 244-249.

- Horstmann, K. T., & Ziegler, M. (2016). Situational perception: its theoretical foundation, assessment, and links to personality. In U. Kumar (Ed.), *The Wiley Handbook Of Personality Assessment* (pp. 31–43). Oxford: Wiley Blackwell.
- Horstmann, K.T., Rauthmann, J. & Sherman, R. (2018). Measurement of situational influences. in V. Zeigler-Hill & T. K. Shackelford. *The Sage Handbook of Personality and Individual Differences* (pp. 465-484). London: SAGE Publications Ltd
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132.
- Horwitz, E. K. (2017). On the misreading of Horwitz, Horwitz, and Cope (1986) and the need to balance anxiety research and the experiences of anxious language learners. In C. Gkonou, M. Daubney & J. M. Dewaele, (Eds.), *New Insights into Language Anxiety: Theory, Research and Educational Implications* (pp. 31–50). Bristol: Multilingual Matters.
- Howard, M. (2005). Second Language Acquisition in a study abroad context: A comparative investigation of the effects of study abroad and foreign language instruction on the L2 learner's grammatical development. *Investigations in instructed Second Language Acquisition. Studies on Language Acquisition*, 25, 495-530.
- Howard, M. (2006). The expression of number and person through verb morphology in advanced French interlanguage. *IRAL – International Review of Applied Linguistics in Language Teaching*, 44(1), 1-22.
- Howard, M. (2008). Morphosyntactic development in the expression of modality: The subjunctive in French L2 acquisition. *Canadian Journal of Applied Linguistics*, 11(3), 171-192.
- Howard, M. (2019). *Study abroad, second language acquisition and interculturality*. Bristol: Multilingual Matters.
- Howard, M., Leme, I., & Regan, V. (2006). The L2 acquisition of a phonological variable: The case of l deletion in French. *Journal of French Language Studies*, 16(1), 1-24.
- Howell, R. T., Ksendzova, M., Nestingen, E., Yerahian, C., & Iyer, R. (2017). Your personality on a good day: How trait and state personality predict daily well-being. *Journal of Research in Personality*, 69, 250-263.
- House of Lords. (2019). *Brexit: the ERASMUS and horizon programmes*. Retrieved March 15, 2020, from <https://publications.parliament.uk/pa/ld201719/ldselect/ldcom/283/283.pdf>
- Houghton, S. A. (2014). Exploring manifestations of curiosity in study abroad as part of intercultural communicative competence. *System*, 42, 368-382.
- Hsieh, G., & Kocielnik, R. (2016). You get who you pay for: The impact of incentives on participation bias. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 823-835).
- Huang, J., & Ryan, A. (2011). Beyond personality traits: A study of personality states and situational contingencies in customer service jobs. *Personnel Psychology*, 64(2), 451-488.
- Huang, T. J., Chi, S. C., & Lawler, J. J. (2005). The relationship between expatriates' personality traits and their adjustment to international assignments. *The International Journal of Human Resource Management*, 16(9), 1656-1670.
- Huensch, A., & Tracy–Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *Modern Language Journal*, 101(2), 275-293.
- Huhn, D., Huber, J., Ippen, F. M., Eckart, W., Junne, F., Zipfel, S., Herzog, W., & Nikendei, C. (2016). International medical students' expectations and worries at the beginning of their medical education: A qualitative focus group study. *BMC Medical Education*, 16-33.
- Hunley, H. A. (2010). Students' functioning while studying abroad: The impact of psychological distress and loneliness. *International Journal of Intercultural Relations*, 34(4), 386-392.

- Husky, M. M., Kovess-Masfety, V., & Swendsen, J. D. (2020). Stress and anxiety among university students in France during Covid-19 mandatory confinement. *Comprehensive Psychiatry*, 152-191.
- Ife, A., Vives Boix, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics*, 4(1), 55-84.
- Ingraham, E. C., & Peterson, D. L. (2004). Assessing the impact of study abroad on student learning at Michigan state university. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10, 83-100.
- Isabelli, C.A. (2001). *The impact of a study abroad experience on the acquisition of L2 Spanish syntax: the null subject parameter* (Unpublished Doctoral Dissertation). University of Illinois, Urbana-Champaign.
- Isabelli-García, C. (2006). Study abroad social networks, motivation, and attitudes: implications for sla. In M. Dufon & E. Churchill (Eds.), *Language Learners in Study Abroad Contexts* (pp. 231–258). Clevedon: Multilingual Matters.
- Isabelli-García, C., Bown, J., Plews, J. L., & Dewey, D. P. (2018). Language learning and study abroad. *Language Teaching*, 51(4), 439-484.
- Jackson, J. (2006). Ethnographic preparation for short-term study and residence in the target culture. *International Journal of Intercultural Relations*, 30(1), 77-98.
- Jackson, J. (2008). Globalization, internationalization, and short-term stays abroad. *International Journal of Intercultural Relations*, 32(4), 349-358.
- Jackson, J. (2010). *Intercultural journeys: From study to residence abroad*. Basingstoke, England; New York: Palgrave Macmillan.
- Jackson, J. (2018). Intervening in the intercultural learning of L2 study abroad students: From research to practice. *Language Teaching*, 51(3), 365-382.
- Jacobone, V. & Moro, G. (2015). Evaluating the impact of the ERASMUS programme: Skills and European identity. *Assessment & Evaluation in Higher Education*, 40(2), 309-328.
- Jochum, C. J. (2014). Measuring the effects of a semester abroad on students' oral proficiency gains: A comparison of at-home and study abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 24, 93-104.
- John, O. P., & Srivastava, S. (1999). The big-five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research*. New York: Guilford Press.
- Johnson, P., & McKinnon, S. (2018). Positioning year-long study abroad at the centre of the modern languages curriculum: Supporting and assessing learning. In J. Plews & K. Misfeldt (Eds.), *Second Language Study Abroad*. Palgrave Macmillan, Cham.
- Josefsson, K., Jokela, M., Cloninger, C. R., Hintsanen, M., Salo, J., Hintsala, T., ... Keltikangas-Järvinen, L. (2013). Maturity and change in personality: Developmental trends of temperament and character in adulthood. *Development and Psychopathology*, 25, 713-727.
- Juan-Garau, M. (2014). Oral accuracy growth after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 87-110). Amsterdam: John Benjamins.
- Juan-Garau, M., & Pérez-Vidal, C. (2007). The effect of context and contact on oral performance in students who go on a stay abroad. *VLAL, Vigo International Journal of Applied Linguistics*, (4), 117-134.
- Juan-Garau, M., Salazar-Noguera, J., & Prieto-Arranz, J. I. (2014). English L2 learners' lexicogrammatical and motivational development at-home and abroad. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 235-258). Amsterdam: John Benjamins Publishing Company.

- Judge, T. A., Simon, L. S., Hurst, C., & Kelley, K. (2014). What I experienced yesterday is who I am today: Relationship of work motivations and behaviors to within-individual variation in the five-factor model of personality. *Journal of Applied Psychology, 99*(2), 199-221.
- Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. P., Breen, W. E., Terhar, D., & Steger, M. F. (2009). The Curiosity and Exploration Inventory-II: Development, factor structure, and psychometrics. *Journal of Research in Personality, 43*(6), 987-998.
- Kim, J., Dewey, D. P., Baker-Smemoe, W., Ring, S., Westover, A., & Eggett, D. L. (2015). L2 development during study abroad in China. *System, 55*, 123-133.
- King, R., & Ruiz-Gelices, E. (2003). International student migration and the European 'year abroad': Effects on European identity and subsequent migration behaviour. *International Journal of Population Geography, 9*(3), 229-252.
- Kinginger, C. (2008). "Language learning in study abroad: Case histories of Americans in France". *Modern Language Journal, 92*, 1-124.
- Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research*. Basingstoke: Palgrave Macmillan
- Kinginger, C. (2011). Enhancing language learning in study abroad. *Annual Review of Applied Linguistics, 31*, 58-73.
- Kinginger, C. (2014). Language socialization in the homestay: American high school students in china. In R. Mitchell, K. McManus & N. Tracy-Ventura (Eds.), *Social Interaction, Identity and Language Learning During Residence Abroad* (pp. 53-74). EUROSLA.
- Kinginger, C. (2015). Student mobility and identity-related language learning. *Intercultural Education, 26*(1), 6-15.
- Klapper, J., & Rees, J. (2003). Reviewing the case for explicit grammar instruction in the university foreign language learning context. *Language Teaching Research, 7*(3), 285-314.
- Klimstra, T. A., & Van Doeselaar, L. (2017). Identity formation in adolescence and young adulthood. In J. Specht (Ed.), *Personality development across the lifespan* (pp. 293-308). London: Elsevier.
- Krashen, S. (1985). *The Input Hypothesis: Issues and implications*. London: Longman.
- Kratochwill, T., & Levin, J. (1978). What time-series designs may have to offer educational researchers. *Contemporary Educational Psychology, 3*(4), 273-329.
- Kwok, O.M., Underhill, A. T., Berry, J. W., Luo, W., Elliott, T. R., & Yoon, M. (2008). Analyzing longitudinal data with multilevel models: An example with individuals living with lower extremity intra-articular fractures. *Rehabilitation Psychology, 53*(3), 370-386.
- Lafford, B., & Collentine, J. G. (2006). The effects of study abroad and classroom contexts on the acquisition of Spanish as a Second language: From research to application. In *the Art of Teaching Spanish: Second Language Acquisition from Research to Praxis* (pp. 103-126). Georgetown University Press.
- Lalonde, R. N., Lee, P. A., & Gardner, R. C. (1987). The common view of the good language learner: An investigation of teachers' beliefs. *Canadian Modern Language Review, 44*(1), 16-34.
- Langley, C. S., & Breese, J. R. (2005). Interacting sojourners: A study of students studying abroad. *The Social Science Journal, 42*(2), 313-321.
- Lara, R., Mora, J. C., & Pérez-Vidal, C. (2015). How long is long enough? L2 English development through study abroad programmes varying in duration. *Innovation in Language Learning and Teaching, 9*(1), 46-57.
- Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. T. Reis (Ed.), *New directions for methodology of social and behavioral sciences* (pp. 41-56). San Francisco: Jossey-Bass.
- Leckie, G. (2019). Cross-classified multilevel models. Retrieved November 3, 2019, from <https://arxiv.org/pdf/1907.02569.pdf>

- Lee, L. (2012). Engaging study abroad students in intercultural learning through blogging and ethnographic interviews. *Foreign Language Annals*, 45(1), 7-21.
- Lee, M. (2015). *The complete history of study abroad*. Retrieved August 11, 2018, from <https://www.gooverseas.com/blog/history-study-abroad>
- Lees, D. (2020). An examination of student experiences of well-being during the year abroad. In S. Salin, D. Hall, & C. Hampton (Eds.), *Perspectives on The Year Abroad: A Selection of Papers from YAC2018* (pp. 11-20). Research-Publishing.Net.
- Lehnart, J., Neyer, F. J., & Eccles, J. (2010). Long-term effects of social investment: The case of partnering in young adulthood. *Journal of Personality*, 78, 639-670.
- Leikas, S., & Salmela-Aro, K. (2015). Personality trait changes among young Finns: the role of life events and transitions. *Journal of Personality*, 83, 117-126.
- Leong, C.H. (2007). Predictive validity of the multicultural personality questionnaire: A longitudinal study on the socio-psychological adaptation of Asian undergraduates who took part in a study-abroad program. *International Journal of Intercultural Relations*, 31(5), 545-559.
- Leong, C. H., & Ward, C. (2000). Identity conflict in sojourners. *International Journal of Intercultural Relations*, 24(6), 763-776.
- Leone, L., Van Der Zee, K. I., Van Oudenhoven, J. P., Perugini, M., & Ercolani, A. P. (2005). The cross-cultural generalizability and validity of the multicultural personality questionnaire. *Personality and Individual Differences*, 38(6), 1449-1462.
- Li, L. (2014). Language proficiency, reading development, and learning context. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 24, 73-92.
- Li, W., Fay, D., Frese, M., Harms, P.D., & Gao, X.Y. (2014). Reciprocal relationship between proactive personality and work characteristics: A latent change score approach. *Journal of Applied Psychology*, 99, 948-965.
- Li, M., Mobley, W., & Kelly, A. (2016). Linking personality to cultural intelligence: An interactive effect of openness and agreeableness. *Personality and Individual Differences*, 89, 105-110.
- Liberati, A, Altman, D, Tetzlaff, J, Mulrow, C, Gøtzsche, C, Ioannidis, P, . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Plos Medicine*, 64-94.
- Lindseth, M. U. (2010). The development of oral proficiency during a semester in Germany. *Foreign Language Annals*, 43(2), 246-268.
- Liu, K. L. (2016). Exploring intercultural competence through an intercultural extracurricular activity in Taiwan. *Journal of Language and Cultural Education*, 4(1), 99-109.
- 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.
- Llanes, À. (2012). The short-and long-term effects of a short study abroad experience: The case of children. *System*, 40(2), 179-190.
- Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37(3), 353-365.
- Llanes, À., & Serrano, R. S. (2011). Length of stay and study abroad: Language gains in two versus three months abroad. *Revista Española De Lingüística Aplicada*, (24), 95-110.
- 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.
- Llanes, À., Tragant, E., & Serrano, R. (2018). Examining the role of learning context and individual differences in gains in L2 writing performance: The case of teenagers on an intensive study-abroad programme. *The Language Learning Journal*, 46(2), 201-216.
- Lodi-Smith, J., & Roberts, B. W. (2007). Social investment and personality: A meta-analysis of the relationship of personality traits to investment in work, family, religion, and volunteerism. *Personality and Social Psychology Review*, 11(1), 68-86.

- Long, M. (1996). The role of the linguistic environment in Second Language Acquisition. In W.C. Ritchie And T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413–468). San Diego, CA: Academic Press.
- López-Serrano, S. (2010). Learning languages in study abroad and at-home contexts: A critical review of comparative studies. *Porta Linguarum: Revista Internacional De Didáctica De Las Lenguas Extranjeras*, (13), 149-164.
- Lucas, R. E. (2018). Re-evaluating the strengths and weaknesses of self-report measures of subjective well-being. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of Well-Being*. Salt Lake City, UT: DEF Publishers.
- Lüdtke, O., Roberts, B., Trautwein, U., & Nagy, G. (2011). A random walk down university avenue: Life paths, life events, and personality trait change at the transition to university life. *Journal of Personality and Social Psychology*, 101(3), 620-637.
- Luhmann, M., Orth, U., Specht, J., Kandler, C., & Lucas, R. (2014). Studying changes in life circumstances and personality: It's about time. *European Journal of Personality*, 28(3), 256-266.
- Macaro, E., Murphy, P., & Vanderplank, R. (2013). Second Language Acquisition: The landscape, the scholarship and the reader. In E. Macaro (Ed.), *The Bloomsbury Companion to Second Language Acquisition* (pp. 3-28). London: Bloomsbury.
- Macintyre, P. D., & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology*, 15, 3-26.
- Macintyre, P. D., & Mercer, S. (2014). Introducing positive psychology to SLA. *Studies in Second Language Learning and Teaching*, 4(2), 153-172.
- Magnan, S. S., & Back, M. (2007). Social interaction and linguistic gain during study abroad. *Foreign Language Annals*, 40(1), 43-61.
- Mahmoodzadeh, M., & Khajavy, G. H. (2019). Towards conceptualizing language learning curiosity in SLA: An empirical study. *Journal of Psycholinguistic Research*, 48(2), 333-351.
- Mallett, R., Hagen-Zanker, J., Slater, R., & Duvendack, M. (2012). The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness: Special Issue on Systematic Reviews*, 4(3), 445-455.
- Marijuan, S., & Sanz, C. (2018). Expanding boundaries: current and new directions in study abroad research and practice. *Foreign Language Annals*, 51(1), 185-204.
- Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory. *British Journal of Clinical Psychology*, 31(3), 301-306.
- Mather, P.C., Karbley, M., & Yamamoto, M. (2012). Identity matters in a short-term, international service-learning program. *Journal of College and Character*, 13(1), 1-14.
- McCrae, R. R., & Costa, P. T. (1990). *Personality in adulthood: A five factor theory perspective* (1st Ed). New York: Guilford Press.
- McCrae, R. R., & Costa, P. T., (1991). Adding Liebe und Arbeit: the full five-factor model and well-being. *Personality and Social Psychology Bulletin*, 17, 227–232.
- McCrae, R. R., & Terracciano, A. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology*, 89(3), 407–425.
- McLeod, M., & Wainwright, P. (2009). Researching the study abroad experience. *Journal of Studies in International Education*, 13(1), 66-71.
- McManus, K., & Mitchell, R. (2015). Subjunctive use and development in L2 French: A longitudinal study. *Language, Interaction and Acquisition*, 6(1), 42-73.
- Meara, P. (1994). "The year abroad and its effects". *Language Learning Journal*, 10, 32-38.
- Mendelson, V. G. (2004). "Hindsight is 20/20:" Student perceptions of language learning and the study abroad experience. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10, 43-63.

- Michl, T., Pegg, K., & Kracen, A. (2019). Gender x culture: A pilot project exploring the study abroad experiences of trans and gender expansive students. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 31(2), 32-50.
- Mikulec, E., Jasper, K., & Cline, L. (2019). "I can navigate the world": Student reflections on study abroad-evidence from one university campus. *Journal for Research and Practice in College Teaching*, 4(1), 84-109.
- Milstein, T. (2005). Transformation abroad: Sojourning and the perceived enhancement of self-efficacy. *International Journal of Intercultural Relations*, 29(2), 217-238.
- Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL. Institut Voor Toegepaste Linguistik*, 17-34.
- Minbashian, A., Beckmann, N., & Wood, R. E. (2018). Emotional intelligence and individual differences in affective processes underlying task-contingent conscientiousness. *Journal of Organizational Behavior*, 39(9), 1182-1196.
- Mischel, W. (1977). On the future of personality measurement. *American Psychologist*, 32(4), 246-254.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102(2), 246-268.
- Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology*, 49, 229-258.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2015). Introduction. In R. Mitchell, N. Tracy-Ventura, & K. McManus (Eds.), *Social Interaction, Identity, and Language Learning During Residence Abroad* (pp. 7-12). Eurosla Monographs.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017a). *Anglophone students abroad: Identity, social relationships, and language learning*. New York: Routledge.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017b). Linguistic development in French. In R. Mitchell, N. Tracy-Ventura, & K. McManus (Eds.), *Anglophone students abroad: Identity, social relationships, and language learning* (pp. 72-104). New York: Routledge.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017c). Linguistic development in Spanish. In R. Mitchell, N. Tracy-Ventura, & K. McManus (Eds.), *Anglophone students abroad: Identity, social relationships, and language learning* (pp. 105-133). New York: Routledge.
- Moher, D., Liberati A, Tetzlaff, J., Altman, G. (2009). The PRISMA group. *Preferred Reporting Items for Systematic Reviews and Meta Analyses: The Prisma Statement*. Plos, Med 6(7).
- Montgomery, J. F., & Arensdorf, J. (2012). Preparing globally competent leaders through innovative study abroad experiences. *Journal of Leadership Studies*, 6(1), 64-71.
- Mora, J. C. (2008). "Learning context effects on the acquisition of a second language phonology". In C. Pérez-Vidal, M. Juan-Garau & A. Bel (Eds.), *A Portrait of The Young in The New Multilingual Spain* (pp. 241-263). Clevedon: Multilingual Matters.
- Mora, J. C. (2014). The role of onset level on L2 perceptual phonological development after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 167-194). Amsterdam: John Benjamins.
- 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.
- Mund, M., & Neyer, F. (2019). Loneliness effects on personality. *International Journal of Behavioral Development*, 43(2), 136-146.
- Murphy-Lejeune, E. (2003). An experience of interculturality: Student travellers abroad. In G. Alred, M. Byram, and M. Fleming (Eds.), *Intercultural Experience and Education* (pp. 101-113). Clevedon, UK: Multilingual Matters

- Natovová, L., & Chýlová, H. (2014). Is there a relationship between self-efficacy, well-being and behavioural markers in managing stress at university students? *Journal on Efficiency and Responsibility in Education and Science*, 7(1), 14-18.
- Neuliep, J. W. (2012). The relationship among intercultural communication apprehension, ethnocentrism, uncertainty reduction, and communication satisfaction during initial intercultural interaction: An extension of anxiety and uncertainty management (AUM) theory. *Journal of Intercultural Communication Research*, 41(1), 1-16.
- Neuliep, J. W., & McCroskey, J. C. (1997). The development of intercultural and interethnic communication apprehension scales. *Communication Research Reports*, 14(2), 145-156.
- Nezlek, J. B. (2001). Multilevel random coefficient analyses of event-and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, 27(7), 771-785.
- Niehoff, E., Petersdotter, L., & Freund, P. A. (2017). International sojourn experience and personality development: Selection and socialization effects of studying abroad and the big five. *Personality and Individual Differences*, 112, 55-61.
- Nieto, N. P., & Nebot, N. (2020). The Cardiff University buddy scheme: how to prepare outgoing students using the experience of the year abroad and final-year students. In S. Salin, D. Hall & C. Hampton (Eds), *Perspectives on the year abroad: A selection of papers from YAC2018* (pp. 43-52). Research-Publishing.Net.
- Noftle, E. E., & Gust, C. J. (2015). powerful situations: some real progress but some future considerations. comment on EJP target article by Rauthmann et al. (2015). *European Journal of Personality*, 29, 404-405.
- Nowell, L., Norris, J., White, D., & Moules, N. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1-13.
- O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, 27(3), 377-402.
- O'Reilly, A., Ryan, D., & Hickey, T. (2010). The psychological well-being and sociocultural adaptation of short-term international students in Ireland. *Journal of College Student Development*, 51(5), 584-598.
- Oberle, E., Ji, X., Magee, C., Guhn, M., Schonert-Reichl, K., & Gadermann, A. (2019). Extracurricular activity profiles and well-being in middle childhood: A population-level study. *Plos ONE*, 14(7), E0218488.
- Oerlemans, W., Bakker, A., & Veenhoven, R. (2011). Finding the key to happy aging: A day reconstruction study of happiness. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 66(6), 665-674.
- Oishi, S., Talhelm, T., & Lee, M. (2015). Personality and geography: Introverts prefer mountains. *Journal of Research in Personality*, 58, 55-68.
- Ożańska-Ponikwia, K. (2010). *Emotions from a bilingual point of view: Perception and expression of emotions in L1 and L2*. (Unpublished Doctoral Dissertation). Birkbeck University, London.
- Ożańska-Ponikwia, K., Dewaele, J. M. (2012). Personality and L2 use: The advantage of being openminded and self-confident in an immigrant context. *EUROSLA Yearbook*, 12, 112-134.
- Ozer, S. (2015). Predictors of international students' psychological and sociocultural adjustment to the context of reception while studying at Aarhus university, Denmark. *Scandinavian Journal of Psychology*, 56(6), 717-725.
- Pampaka, M., Hutcherson, G., & Williams, J. (2016). Handling missing data: Analysis of a challenging data set using multiple imputation. *International Journal of Research & Method in Education*, 39(1), 19-37.

- Parker, P., Lüdtke, O., Trautwein, U., & Roberts, B. (2012). Personality and relationship quality during the transition from high school to early adulthood. *Journal of Personality*, 80(4), 1061-1089.
- Parrigon, S., Woo, S., Tay, L., Wang, T., & Cooper, M. Lynne. (2017). CAPTION-ing the situation: A lexically derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, 112(4), 642-681.
- Peifer, J., & Yangchen, T. (2017). Exploring cultural identity, personality, and social exposure correlates to college women's intercultural competence. *SAGE Open Journals*, 7(2), 1-13.
- Peugh, J. L. (2010). A practical guide to multilevel modelling. *Journal of School Psychology*, 48(1), 85-112.
- Pellegrino, V. A. (1998). Student perspectives on language learning in a study abroad context. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 4(1), 91-120.
- Pelletier, S. G. (2019). Refraining the re-entry experience. *International Educator*, 28(3), 50-51.
- Pérez-Vidal, C. (2014). Study abroad and formal instruction contrasted. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 17– 58). Amsterdam: John Benjamins.
- Pérez-Vidal, C., & Juan-Garau, M. (2009). The effect of study abroad (SA) on written performance. *EUROSLA Yearbook*, 9(1), 269-295.
- Pérez-Vidal, C., & Juan-Garau, M. (2011). The effect of context and input conditions on oral and written development: A study abroad perspective. *International Review of Applied Linguistics in Language Teaching*, 49(2), 157-185.
- Pérez-Vidal, C., & Barquin, E. (2014). Comparing progress in academic writing after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 217-234). Amsterdam: John Benjamins.
- Petersen, S., Divitini, M., & Chabert, G. (2008). Identity, sense of community and connectedness in a community of mobile language learners. *Recall*, 20(3), 361-379.
- Pituch, K. A., & Stevens, J.P. (2016). *Applied multivariate statistics for social sciences* (6th Ed.). New York: Routledge.
- Pinar, A. (2018). Learning languages abroad: The influence of the length of stay on communicative competence. In *Study Abroad Contexts for Enhanced Foreign Language Learning* (pp. 23-43). IGI Global.
- Potter, H. (2020). On the year abroad experience and positive well-being. In S. Salin, D. Hall & C. Hampton (Eds), *Perspectives on the year abroad: A selection of papers from YAC2018* (pp. 21-29). Research-Publishing.Net.
- Poyrazli, S., & Mitchell, M. A. (2020). Mental health problems of US students studying abroad. *Journal of International Students*, 10(1), 17-27.
- Puppel, S. (2012). Remarks on learner autonomy and success in foreign language learning as a measure of learner resilience. *An International Journal of Applied Linguistics*, 39(1), 7-12.
- Raatz, U., & Klein-Braley, C. (1981). The C-Test – A modification of the cloze procedure. In T. Culhane, C. Klein-Braley & D. K. Stevenson (Eds.), *Practice and Problems in Language Testing IV* (pp. 113-138). Colchester: University of Essex.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212.
- Ranta, L., & Meckelborg, A. (2013). How much exposure to English do international graduate students really get? Measuring language use in a naturalistic setting. *Canadian Modern Language Review*, 69(1), 1-33.
- Rauthmann, J., Gallardo-Pujol, D., Guillaume, E., Todd, E., Nave, C., Sherman, R., . . . Funder, D. (2014). The situational eight diamonds: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, 107(4), 677-718.

- Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality, 29*(3), 363-381.
- Rauthmann, J., & Sherman, R. (2016). Ultra-brief measures for the Situational Eight Diamonds domains. *European Journal of Psychological Assessment, 32*(2), 165-174.
- Rees, J., & Klapper, J. (2007). Analysing and evaluating the linguistic benefit of residence abroad for UK foreign language students. *Assessment and Evaluation in Higher Education, 32*(3), 331-353.
- Regan, V. (1995). The acquisition of sociolinguistic native speech norms. In B.F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 245-267), Amsterdam: John Benjamins.
- Richter, J., Zimmermann, J., Neyer, F. J., & Kandler, C. (2020). Do sojourn effects on personality trait changes last? A five-year longitudinal study. *European Journal of Personality, 1-22*.
- Roberts, B. W., & Wood, D. (2006). personality development in the context of the neo-socioanalytic model of personality. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of Personality Development* (pp. 11-39). New York: Lawrence Erlbaum Associates Publishers.
- Roberts, B. W., & Robins, R. W. (2004). A longitudinal study of person-environment fit and personality development. *Journal of Personality, 72*(89), 89-110.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006a). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*(1), 1-25.
- Roberts, B. W., Harms, P. D., Smith, J., Wood, D., & Webb, M. (2006b). Methods in personality psychology. In M. Eid & E. Diener (Eds.), *Handbook of Psychological Assessment: A Multimethod Perspective* (pp. 321-335). Washington, DC: American Psychological Association.
- Roberts, B. W., & Nickels, L. B. (2017). A critical evaluation of the neo-socioanalytic model of personality. In Specht, J. (Ed.), *Personality development across the lifespan* (pp. 157-178). London: Elsevier.
- Roberts, C., Byram, M., Barro, A., Jordan, S., & Street, B. (2001). *Language learners as ethnographers*. Clevedon: Multilingual Matters.
- Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality, 69*(4), 617-640.
- Robinson, D., Gabriel, N., & Katchan, O. (1994). Personality and second language learning. *Personality and Individual Differences, 16*(1), 143-157.
- Roli, G., Monari, P. (2014). A review of multilevel modelling: Some methodological issues and advances. In V. Melas V, S. Mignani, P. Monari, L. Salmaso (Eds.), *Topics in Statistical Simulation* (pp. 431-438). New York: Springer.
- Roos, A. L., Goetz, T., Voracek, M., Krannich, M., Bieg, M., Jarrell, A., & Pekrun, R. (2020). Test anxiety and physiological arousal: A systematic review and meta-analysis. *Educational Psychology Review, 1-40*.
- Rosenman, R., Tennekoon, V., & Hill, L. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research, 2*(4), 320-332.
- Røysamb, E., Nes, R. B., Czajkowski, N. O., & Vassend, O. (2018). Genetics, personality and well-being. a twin study of traits, facets and life satisfaction. *Scientific Reports, 8*(1), 1-13.
- Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly, 9*(1), 41-51.
- Russell, J., Thomson, G., & Rosenthal, D. (2008). International student use of university health and counselling services. *Higher Education, 56*(1), 59-75.
- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52*(1), 141-166.
- Salkind, N. J. (2010). *Encyclopaedia of research design*. Thousand Oaks, CA: SAGE Publications.
- Salgado, J., Moscoso, S., Sanchez, J., Alonso, P., Choragwicka, B., & Berges, A. (2015). Validity of the Five-factor Model and their facets: The impact of performance measure and facet

- residualization on the bandwidth-fidelity dilemma. *European Journal of Work and Organizational Psychology*, 24(3), 325-349.
- Sam, D. L. (2001). Satisfaction with life among international students: an exploratory study. *Social Indicators Research*, 53(3), 315-337.
- Sanz, C. (2014). Contributions of study abroad research to our understanding of SLA processes and outcomes: the SALA project, an appraisal. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 1-16). Amsterdam: John Benjamins.
- Sanz, C. & Morales-Front, A. (2018). Issues in study abroad research and practice. In C. Sanz & A. Morales-Front (Eds.), *The Routledge Handbook of Study Abroad Research and Practice* (pp. 1-15). New York: Routledge.
- Savicki, V., Arrúe, C., & Binder, F. (2013). Language fluency and study abroad adaptation. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 22(1), 37-57.
- Savicki, V., & Price, M. V. (2017). Components of reflection: A longitudinal analysis of study abroad student blog posts. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 29(2), 51-62.
- Shaftel, J., Shaftel, T., & Ahluwalia, R. (2007). International educational experience and intercultural competence. *International Journal of Business & Economics*, 6(1), 25-34.
- Schartner, A. (2016). The effect of study abroad on intercultural competence: A longitudinal case study of international postgraduate students at a British university. *Journal of Multilingual and Multicultural Development*, 37(4), 402-418.
- Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, 31(3), 699-714.
- Schmidt-Rinehart, B. C., & Knight, S. M. (2004). The homestay component of study abroad: Three perspectives. *Foreign Language Annals*, 37(2), 254-262.
- Schmutte, P. S., & Ryff, C. D. (1997). Personality and well-being: Re-examining methods and meanings. *Journal of Personality and Social Psychology*, 73(3), 549-559.
- Schumann, J. H. (1978). *The pidginization process: A model for second language acquisition*. Rowley, MA: Newbury House.
- Schurer, S., Kassenboehmer, S., & Leung, F. (2015). Do universities shape their students' personality? *IZA Discussion Papers*, 8873, Institute for The Study of Labor (IZA), Bonn.
- Selphout, M., Burk, W., Branje, S., Denissen, J., Van Aken, M., & Meeus, W. (2010). Emerging late adolescent friendship networks and big five personality traits: A social network approach. *Journal of Personality*, 78, 509-538.
- Severiens, S. E., & Schmidt, H. G. (2009). Academic and social integration and study progress in problem-based learning. *Higher Education*, 58(1), 59-69.
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *British Medical Journal*, 349-374.
- Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology*, 109(5), 872- 888.
- Shi, J., Yao, Y., Zhan, C., Mao, Z., Yin, F., & Zhao, X. (2018). The relationship between big five personality traits and psychotic experience in a large non-clinical youth sample: The mediating role of emotion regulation. *Frontiers in Psychiatry*, 9, 648.
- Shively, R. L. (2011). L2 pragmatic development in study abroad: A longitudinal study of Spanish service encounters. *Journal of Pragmatics*, 43, 1818-1835.
- Shively, R. L. (2013). Learning to be funny in Spanish during study abroad: L2 humor development. *The Modern Language Journal*, 97(4), 930-946.
- Searle, W., & Ward, C. (1990). The prediction of psychological and sociocultural adjustment during cross-cultural transitions. *International Journal of Intercultural Relations*, 14(4), 449-464.

- Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at-home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173-199.
- Sellars, V. (2000). Anxiety and reading comprehension in Spanish as a foreign language. *Foreign Language Annals*, 33(5), 512-521.
- Serrano, R., Llanes, &, & 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.
- 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.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shiner, R. L., & Deyoung, C. G. (2013). The structure of temperament and personality traits: A developmental perspective. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology* (pp. 113–141). New York, Oxford University Press.
- Sigalas, E. (2010). Cross-border mobility and European identity: The effectiveness of intergroup contact during the ERASMUS year abroad. *European Union Politics*, 11(2), 241-265.
- Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, 15(3), 194-200.
- Smith, R. A. & Khawaja, N. G. (2011). A review of the acculturation experiences of international students. *International Journal of Intercultural Relations*, 35(6), 699–713.
- Smith, B. P., & Yang, W. (2017). Learning outcomes in an interdisciplinary study abroad program: Developing a global perspective. *Journal of Family and Consumer Sciences*, 109(1), 43-50.
- Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, 101, 862–882.
- Specht, J. (2017). *personality development across the lifespan*. London: Academic Press.
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (form y)*. Palo Alto, CA: Consulting Psychologists Press.
- Soria, K. M., & Troisi, J. (2014). Internationalization at-home alternatives to study abroad: Implications for students' development of global, international, and intercultural competencies. *Journal of Studies in International Education*, 18(3), 261-280.
- Soto, C. (2015). Is happiness good for your personality? Concurrent and prospective relations of the Big Five with subjective well-being. *Journal of Personality*, 83(1), 45-55.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, 94(4), 718–737.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, 100(2), 330-348.
- Soto, C. J., Kronauer, A., & Liang, J. K. (2016). Five-factor model of personality. In S. K. Whitbourne (Ed.), *Encyclopedia of adulthood and aging* (pp. 506-510). Hoboken, NJ: Wiley.
- Souto-Otero, M., Huisman, J., Beerkens, M., De Wit, H., & Vujić, S. (2013). Barriers to international student mobility: Evidence from the ERASMUS program. *Educational Researcher*, 42(2), 70-77.
- Stronkhorst, R. (2005). Learning outcomes of international mobility at two Dutch institutions of higher education. *Journal of Studies in International Education*, 9(4), 292–315.

- Stamp, E., Crust, L., Swann, C., Perry, J., Clough, P., & Marchant, D. (2015). Relationships between mental toughness and psychological well-being in undergraduate students. *Personality and Individual Differences*, 75, 170-174.
- Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, 134(1), 138-161.
- Stephan, Y., Sutin, A. R., & Terracciano, A. (2014). Physical activity and personality development across adulthood and old age: Evidence from two longitudinal studies. *Journal of Research in Personality*, 49, 1-7.
- Stephan, Y., Boiché, J., Canada, B., & Terracciano, A. (2014). Association of personality with physical, social, and mental activities across the lifespan: Findings from us and French samples. *British Journal of Psychology*, 105(4), 564-580.
- Stewart-Brown, S., Evans, J., Patterson, J., Petersen, S., Doll, H., Balding, J., & Regis, D. (2000). The health of students in institutes of higher education: An important and neglected public health problem? *Journal of Public Health*, 22(4), 492-499.
- Street, H., Nathan, P., Durkin, K., Morling, J., Dzahari, M. A., Carson, J., & Durkin, E. (2004). Understanding the relationships between well-being, goal setting and depression in children. *Australian & New Zealand Journal of Psychiatry*, 38(3), 155-161.
- Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, 18(2), 103-118.
- Suanet, I., & Van De Vijver, F. J. (2009). Perceived cultural distance and acculturation among exchange students in Russia. *Journal of Community & Applied Social Psychology*, 19(3), 182-197.
- Sullivan, G., & Feinn, R. (2012). Using effect size-or why the *p*-value is not enough. *Journal of Graduate Medical Education*, 4(3), 279-82.
- Swain, M. (1985). Communicative competence: Some roles of Comprehensible Input and Comprehensible Output in its development. In S. Gass And C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Taguchi, N. (2015). Cross-cultural adaptability and development of speech act production in study abroad. *International Journal of Applied Linguistics*, 25(3), 343-365.
- 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.
- Tao, K. M., Li, X. Q., Zhou, Q. H., Moher, D., Ling, C. Q., & Yu, W. F. (2011). from QUOROM to PRISMA: A survey of high-impact medical journals' instructions to authors and a review of systematic reviews in anaesthesia literature. *Plos One*, 6(11), 1-5.
- Teichler, U. (1997). *The ERASMUS experience. Major findings of the ERASMUS evaluation research*. Luxembourg: Office for Official Publications of the European Communities.
- Teichler, U. (2004). Temporary study abroad: The life of ERASMUS students. *European Journal of Education*, 39(4), 395-408.
- Teichler, U. (2015). The impact of temporary study abroad. In R. Mitchell, K. McManus & N. Tracy-Ventura (Eds.), *Social Interaction, Identity and Language Learning During Residence Abroad* (pp. 15-32). EUROSLA.
- Teichler, U., & Janson, K. (2007). The professional value of temporary study in another European country: Employment and work of former ERASMUS students. *Journal of Studies in International Education*, 11(3-4), 486-495.
- Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, 41(2), 363-387.
- Teppers, E., Klimstra, T. A., Van Damme, C., Luyckx, K., Vanhalst, J., & Goossens, L. (2013). Personality traits, loneliness, and attitudes towards aloneness in adolescence. *Journal of Social and Personal Relationships*, 30, 1045-1063.

- Thomas, D. (2018). Factors that contribute to homesickness among students in Thailand. *Kasetsart Journal of Social Sciences*, 1-6.
- Thomas, J., Brunton, J., Graziosi, S. (2010). *EPPI-Reviewer 4: Software for research synthesis*. EPPI-Centre Software. London: Social Science Research Unit, UCL Institute of Education.
- Tkach, C., & Lyubomirsky, S. (2006). How do people pursue happiness? Relating personality, happiness-increasing strategies, and well-being. *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being*, 7(2), 183–225.
- Torgerson, C. (2003). *Systematic reviews*. London: Continuum.
- Torgerson, D. J., & Torgerson, C. J. (2003). Avoiding bias in randomised controlled trials in educational research. *British Journal of Educational Studies*, 51(1), 36-45.
- Torgerson, C., Hall, J., & Light, K. (2012). Systematic reviews. In J. Arthur, M. Warring, R. Coe, & L. Hedges (Eds.), *Research Methods and Methodologies in Education* (pp. 217-230). London: Sage.
- 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.
- 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.
- Tragant, E. (2012). Change or stability in learners’ perceptions as a result of study abroad. In C. Munoz (Ed.), *Intensive Exposure Experiences in Second Language Learning* (pp. 161- 190). Bristol: Multilingual Matters.
- Trenchs-Parera, M. (2009). Effects of formal instruction and a stay abroad on the acquisition of native-like oral fluency. *Canadian Modern Language Review*, 65(3), 365-393.
- Trentman, E. (2013). Imagined communities and language learning during study abroad: Arabic learners in Egypt. *Foreign Language Annals*, 46, 545–564.
- Tulloch, B., & Ortega, L. (2018). Fluency and multilingualism in study abroad: Lessons from a scoping review. *System*, 71, 7-21.
- Twisk, J., & De Wente, W. (2002). Attrition in longitudinal studies: How to deal with missing data. *Journal of Clinical Epidemiology*, 55(4), 329-337.
- UMHAN. (2013). *Considering the needs of students with a mental health condition*. Retrieved March 28, 2018, from https://hubble-live-assets.s3.amazonaws.com/umhan/attachment/file/13/resource_for_study_needs_assessors_-_mental_health_conditions_october_2013.pdf
- University of Liverpool. (2020). *Study abroad buddies*. Retrieved June 10, 2020, from <https://www.liverpool.ac.uk/study-abroad/buddies/>
- Universities UK International. (2019). *Gone international: Rising aspirations*. Retrieved July 10, 2019, from <https://www.universitiesuk.ac.uk/international/documents/2019/gone-intl-2019.pdf>
- Valls-Ferrer, M. & Mora, J. C. (2014). The role of onset level on L2 perceptual phonological development after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 111-136). Amsterdam: John Benjamins.
- Van Vianen, A. E. M. & De Dreu, C. K. W. (2001). Personality in teams: Its relations to social cohesion, task cohesion, and team performance. *European Journal of Work and Organizational Psychology*, 10, 97– 120.
- Vande Berg, M., Connor-Linton, J., & Paige, R. M (2009). The Georgetown consortium project: Interventions for student learning abroad. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 18, 1-75.
- Van Daele, S., Housen, A., Pierrard, M., & Debruyne, L. (2006). The effect of extraversion on oral L2 proficiency. *EUROSLA Yearbook*, 6, 213–236.

- Van Der Zee, K. I., & Van Oudenhoven, J. P. (2000). The multicultural personality questionnaire: A multidimensional instrument of multicultural effectiveness. *European Journal of Personality, 14*(4), 291-309.
- Van Der Zee, K., Van Oudenhoven, J. P., Ponterotto, J. G., & Fietzer, A. W. (2013). Multicultural personality questionnaire: Development of a short form. *Journal of Personality Assessment, 95*(1), 118-124.
- Vanhalst, J., Klimstra, T. A., Luyckx, K., Scholte, R. H. J., Engels, R. C. M. E., & Goossens, L. (2012). The interplay of loneliness and depressive symptoms across adolescence: Exploring the role of personality traits. *Journal of Youth and Adolescence, 41*, 776-787.
- Van Niejenhuis, C., Otten, S., & Flache, A. (2018). Sojourners' second language learning and integration. The moderating effect of multicultural personality traits. *International Journal of Intercultural Relations, 63*, 68-79.
- Verhoeven, L., & Vermeer, A. (2002). Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics, 23*(3), 361-374.
- Vittersø, J., & Nilsen, F. (2002). The conceptual and relational structure of subjective well-being, neuroticism, and extraversion: Once again, neuroticism is the important predictor of happiness. *Social Indicators Research, 57*(1), 89-118.
- Von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science, 6*(6), 574-588.
- Walker, C., & Braskamp, L. A. (2008). *Relationship between well being and global perspective*. Chicago: Global Perspective Institute.
- Ward, C., & Chang, W. C. (1997). "Cultural fit": A new perspective on personality and sojourner adjustment. *International Journal of Intercultural Relations, 21*, 525-533.
- Ward, C., Leong, C., & Low, M. (2004). Personality and sojourner adjustment: An exploration of the big five and the cultural fit proposition. *Journal of Cross-Cultural Psychology, 35*, 137-151.
- Warga, M., & Schölmlberger, U. (2007). The acquisition of French apologetic behavior in a study abroad context. *Intercultural Pragmatics, 4*(2), 221-251.
- Wawera, A. S., & McCamley, A. (2019). Loneliness among international students in the UK. *Journal of Further and Higher Education, 1*-13.
- Weatherley, D. (1964). Self-perceived rate of physical maturation and personality in late adolescence. *Child Development, 11*97-1210.
- Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality. Special Issue: Personality and Daily Experience, 59*(3), 339-354.
- Wielkiewicz, R. M., & Turkowski, L. W. (2010). Re-entry issues upon returning from study abroad programmes. *Journal of College Student Development, 51*(6), 649-664.
- Wilkinson, S. (1998). Study abroad from the participants' perspective: a challenge to common beliefs. *Foreign Language Annals, 31*(1), 23-39.
- Wilkinson, S. (2002). The omnipresent classroom during summer study abroad: American students in conversation with their French hosts. *Modern Language Journal, 86*, 157-173.
- Wilkinson, T.J., & Hansen, J. C. (2006). The relationship among leisure interests, personality traits, affect, and mood. *Measurement & Evaluation in Counselling & Development, 39*, 31-41.
- Wiseman, H. (1997). Far away from home: The loneliness experience of overseas students. *Journal of Social and Clinical Psychology, 16*(3), 277-298.
- Williams, T.R. (2005). Exploring the impact of study abroad on students' intercultural communication skills: Adaptability and sensitivity. *Journal of Studies in International Education, 9*(4), 356-371.
- Willis, F.M., Doble, G., Sankarayya, U., & Smithers, A. (1977). *Residence abroad and the student of modern languages: A preliminary survey*. Bradford: University of Bradford Modern Languages Centre.

- Wilson, R. (2008). *“Another language is another soul”: individual differences in the presentation of self in a foreign language* (unpublished doctoral dissertation). University of Birkbeck, London.
- Wilt, J., Nofle, E., Fleeson, W., & Spain, J. (2012). The dynamic role of personality states in mediating the relationship between extraversion and positive affect. *Journal of Personality, 80*(5), 1205-1236.
- Winke, P. (2017). Using focus groups to investigate study abroad theories and practice. *System, 71*, 73-83.
- Winke, P., Gass, S., & Heidrich, E. (2019). Modern-day foreign language majors: Their goals, attainment, and fit within a twenty-first century curriculum. In P. Winke & S. Gass (Eds.), *Foreign Language Proficiency in Higher Education* (pp. 117-136). Berlin/New York: Springer.
- Wodak, R. (2017, September). *Discourse, politics, identity: The discursive construction of Austrian national identities 1995-2015*. Presentation presented at the university of Vienna, Vienna.
- Wolff, F., Schmidt, F., Borzиковsky, C., Möller, J., & Wagner, J. (2020). Educational stays abroad and the development of self-perceived grit: a longitudinal analysis in young adulthood. *Current Psychology, 1-16*.
- Wood, D., Crapnell, T., Lau, L., Bennett, A., Lotstein, D., ... & Kuo, A. (2018). Emerging adulthood as a critical stage in the life course. In N. Halfon, C. Forrest, R. Lerner, & E. Faustman (Eds.), *Handbook of Life Course Health Development* (pp. 123-143). Cham: Springer.
- Wood, R., Beckmann, N., Birney, D. P., Beckmann, J. F., Minbashian, A., & Chau, R. (2019). Situation contingent units of personality at work. *Personality and Individual Differences, 136*, 113-121.
- Wu, H., & Zhang, L. J. (2017). Effects of different language environments on Chinese graduate students' perceptions of English writing and their writing performance. *System, 65*, 164-173.
- Yakunnina, E., Weigold, I. & McCarthy, A. (2011). Group counselling with international students: Practical, ethical and cultural considerations. *Journal of College Student Psychotherapy, 25*(1), 67-78.
- Yamagata, S., Suzuki, A., Ando, J., Ono, Y., Kijima, N., Yoshimura, K., . . . Jang, K. (2006). Is the genetic structure of human personality universal? A cross-cultural twin study from North America, Europe, and Asia. *Journal of Personality and Social Psychology, 90*(6), 987-998.
- Yang, J. S. (2016). The effectiveness of study-abroad on second language learning: A meta-analysis. *Canadian Modern Language Review, 72*(1), 66-94.
- Yang, Z., Asbury, K., & Griffiths, M. D. (2019). An exploration of problematic smartphone use among Chinese university students: Associations with academic anxiety, academic procrastination, self-regulation and subjective well-being. *International Journal of Mental Health and Addiction, 17*(3), 596-614.
- Yen, W. J., & Stevens, P. (2004). Taiwanese students' perspectives on their educational experiences in the United States. *International Education Journal, 5*(3), 294-307.
- Xu, W., Case, R. E., & Wang, Y. (2009). Pragmatic and grammatical competence, length of residence, and overall L2 proficiency. *System, 37*, 205-216.
- Zhang, J., Beckmann, N., & Beckmann, J. F. (2019). One situation doesn't fit all: variability and stability of state willingness to communicate in a Chinese college English classroom. *Language Teaching Research, 1-26*.
- Ziegler, M., Horstmann, K. T., & Ziegler, J. (2019). Personality in situations: Going beyond the ocean and introducing the situation five. *Psychological Assessment, 31*(4), 567-580.
- Zimmermann, J., & Neyer, F. (2013). Do we become a different person when hitting the road? personality development of sojourners. *Journal of Personality and Social Psychology, 105*(3), 515-530.

Appendices

Appendix A: Systematic Review Protocol

Title:

A systematic review of the impact of participation in ERASMUS and other study abroad programmes on linguistic competency development.

Introduction

Rationale

There appears public consensus that for foreign language undergraduate students, language development during a period of study abroad far outweighs that during a period of study remaining at-home (Hessel, 2017). Reasons for this stem from the assumed linguistic opportunities provided and the willingness of these learners to respond accordingly. Such a notion is supported by Second Language Acquisition theories such as the Output Hypothesis (Swain, 1985) and the Interaction Hypothesis (Long, 1996), which sees sojourning as the superior context for it allows optimal language input, output and consequently uptake. However, contrary to such expectations, the literature available provides an ambiguous and at times contradicting picture, with these gains often being subtle and subject to both learner-internal and learner-external factors (e.g., Collentine, 2009; Marijuan & Sanz, 2018). These can include personality (e.g., Baker-Smemoe, Dewey, Brown, & Martinsen, 2014), starting proficiency level (e.g., Golonka, 2006) and attitude/motivation (e.g., Gardner, 1985), while learner external-factors centre on the learning environment. For example, host country native speakers may not wish to correct learners. Living accommodation and geographical location (i.e., dialect and slang) can all further inhibit or facilitate the extent of one's linguistic growth (e.g., DeKeyser, 2007; Coleman, 2015; Jackson, 2018). Empirical evidence further suggests that learning context (home vs abroad) can influence both type and rate of skill development. Those abroad are more likely to develop in the skills of speaking and listening, while those who remain at-home improve most in writing and reading (Llanes, Tragant, & Serrano, 2018). This dichotomy is likely a reflection on skill use and practice. Emphasis in the classroom is often on reading and writing development, while abroad, learners actively use the language to communicate daily.

Anecdotal evidence collected through research and organisational questionnaires continues to present the superior nature of the sojourning context. In a report by the British Council (2015), 82% of the 682 language learners questioned saw improving their language skills as a significant motivator in completing a year abroad. 491 of these reported back that they perceived being abroad as having a substantial impact on their linguistic ability. While these reports are descriptive, they failed to ascertain in which areas perceived growth was experienced, while self-report data is further inhibited by inflation and the notion of the 'ideal self'. This mismatch is some ways troubling if we are to provide sojourners with realistic expectations of language development abroad, and this review can be used to synthesise the literature in, for example, a workshop environment.

As a field, the literature has built its foundations through American exchange programmes. Interest from publishers such as Frontiers and Routledge has fuelled the topic interest but has led some to criticise the imbalance in studies between that of American and European origin

(e.g., Rees & Klapper, 2007; Llanes, 2011). In turn, this review wishes to place interest in studies investigating the ERASMUS exchange programme and those programmes of which are akin but based outside of Europe. This balanced approach aims to provide a fair comparison between exchange programme type and provide a synthesis of the possible variables (e.g., duration of program) which can account for the differences in linguistic change witnessed. There is currently no review, systematic or otherwise, which has focused on ERASMUS centred studies as a subset of all other exchange programmes worldwide.

In light of contradictory evidence, it's important that make available a summary of research concerning language change in order to ensure sojourners have realistic expectations of the year abroad as a whole. Past research (e.g., Mendelsohn, 2004; Badstübner & Ecke, 2009) has continually shown students to over-inflate their language gain expectations. Here, perceived post-test language growth scores are significantly lower than those believed at pre-test. These studies have demonstrated post-test scores to be far lower than those hypothesised in the pre-test. In instances where expectations are not met, these experiences are likely to result in disappointment (e.g., Stroebe, Lenkert, & Jonas, 1988; Wilkinson, 1998).

Objectives

The aim of this systematic review is to assess and synthesise the impact of study abroad programmes worldwide on sojourners' linguistic ability compared to non-sojourners who remain in the domestic classroom. As such, the research question is as follows:

- 1) How effective are study abroad programmes in achieving their fundamental aim of improving linguistic ability in all four skills compared to those who remain in the domestic classroom?

Methods

Study design

This systematic review will be conducted according to methodology and standards developed by the Campbell Collaboration and the results presented according to the PRISMA statement. This protocol was also developed using the PRISMA-P guidelines and checklist.

Eligibility criteria

Due to the imbalance in the literature available (see rationale), the study design criteria will be slightly different according to the exchange programme investigated.

Study designs (non-ERASMUS studies)

Due to the greater number of studies which investigate study abroad programmes worldwide only those studies which have a control group will be eligible. As such, Randomised Controlled Trials and Quasi-experimental designs with a control group will be included. Any pre-experimental design (pre/post) will be excluded. This decision was made in order to derive a stronger causal inference on the efficacy of the study abroad to yield linguistic change. Without a control group, this inference is weakened, and fewer rigorous conclusions can be drawn.

Study designs (non-ERASMUS studies)

Included Designs	Excluded Designs
Randomised controlled trials (individual and cluster) Regression-discontinuity design	
Quasi-experimental designs: <ul style="list-style-type: none"> • a non-equivalent control group post-test-only • non-equivalent control group pre-test/post-test Time series design: <ul style="list-style-type: none"> • interrupted time series design • control time-series design with equivalent or non-equivalent control group 	Quasi-experimental designs: <ul style="list-style-type: none"> • one-group post-test only • one-group post-test only design with multiple substantive post-tests • basic time series designs
	Ex post facto control group designs Pre-experimental design (pre-post or before and after) Observational design

Study designs (ERASMUS studies)

Studies which are ERASMUS centred will be included if they have a comparison group or are pre/post-test experiment design. This difference increases inclusion sensitivity and will ensure that all relevant studies are included. The findings from this aspect of the review hold greater external validity for the UK context, and as such, a lower threshold for rigour in deriving causal inference will be allowed. There is a strong possibility that without such concessions, very few papers will be included. As noted in the rationale, this will to the best of the author's knowledge be the first systematic review which synthesises the ERASMUS literature. It is consequently important that as many suitable papers are included. All other inclusion/exclusion criteria seen above will remain the same for ERASMUS studies.

Included Designs	Excluded Designs
Randomised controlled trials (individual and cluster) Regression-discontinuity design	
Quasi-experimental designs: <ul style="list-style-type: none"> • a non-equivalent control group post-test-only • non-equivalent control group pre-test/post-test Time series design: <ul style="list-style-type: none"> • interrupted time series design • control time-series design with equivalent or non-equivalent control group 	Quasi-experimental designs: <ul style="list-style-type: none"> • one-group post-test only • one-group post-test only design with multiple substantive post-tests • basic time series designs
Pre-experimental design (pre-post or before and after) Observational design	Ex post facto control group designs

Participants

The participants included in this review will be either undertaking a period abroad as part of their academic degree programme or remaining home in the domestic classroom. The individuals will

more than likely be language bachelor students who are undertaking a period abroad in order to develop their linguistic proficiency. Participants can be undertaking either study or work whilst abroad. Those partaking in activities with affiliated organisations (e.g., teaching with the British Council) will also be included. Where applicable, the control group will be formed of domestic students of whom are based in the home country of the treatment group (e.g., L1 (native) Spanish sojourners in Germany will be compared to L1 Spanish domestic students (remaining in Spain).

Intervention

Of interest to this review, is any study which includes study abroad as the intervention with the objective of improving one's linguistic ability. The intervention (study abroad) is a naturally occurring, non-manipulated intervention. Any intervention which does not meet the highlighted criteria and/or does not have a comparison/control group will be excluded. In sum, the intervention programme must 5+ to 1 year in length (summer schools are excluded) and include learners who are partaking in the intervention as part of their academic studies.

	ERASMUS intervention criteria
General Characteristics	<ul style="list-style-type: none"> • Serves as a mobility programme during an individual's degree programme. • Typically sends students from one European country to another. • Focuses on studying (i.e., going to another university) or completing an ERASMUS+ traineeship.
Participants	<ul style="list-style-type: none"> • Open to students at any higher education institution that has an ERASMUS University Charter. Students do not have to be learning a language to partake, but many who do go abroad for language training are prior language learners. • Can also include Master students under the ERASMUS Mundus Joint Master Degrees.
Length	<ul style="list-style-type: none"> • Typically 3 – 12 months. Summer Schools are excluded.
Countries	<ul style="list-style-type: none"> • All 28 EU countries participate and non-EU countries of the Republic of Macedonia, Iceland, Norway, Liechtenstein and Turkey. • Albania, Algeria, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Egypt, Georgia, Israel, Jordan, Kosovo, Lebanon, Libya, Moldova, Montenegro, Morocco, Palestine, Russia, Serbia, Syria, Switzerland, Tunisia and Ukraine can also participate in most of the features of ERASMUS including mobility.

Timing

There is no time restriction, although one should be aware of how the nature of study abroad programmes have changed over the past decades (Donatelli, 2010). With the onset of the internet and cheaper travel, study abroad programmes are not as immersive as they once were.

Setting

The population selected can be undertaking any type of study abroad as long it meets the criteria set out above, both in intervention and study design. Control students should be in the home country of the sojourning students in order for fair comparison (e.g., L1 Spanish sojourners in Germany will be compared to L1 Spanish domestic students). There is no restriction on setting context (i.e., placement, teaching, studying). Please note that different contexts may have different titles in journals, e.g., the British Council is the organisation which deals with teaching contexts but is an affiliated partner.

Publication Status

In order to avoid any publication bias, both published and non-published literature will be included. Research included in the review includes in the form of journal articles; book chapters; theses etc.

Language

There is no language bias. Studies of any language will be included, and where required, a native speaker of the particular language will be consulted a native speaker.

Information Sources

Literature searches will be conducted using the Durham Online Database tool (found at <https://www.dur.ac.uk/library/resources/online/databases/>) under the topics of Education and Psychology. The databases scanned although not limited to will include Web of Science; ERIC; British Education Index and PyscARTICLES. The literature search will be limited to the English language.

To ensure saturation, reference lists will be scanned to ensure the inclusion of any other possible relevant studies not picked up through the databases. Moreover, if other relevant systematic reviews and/or meta-analyses are found, these will also be scanned for any relevant studies.

Search strategy

Web of Science (Social Science Citation Index & Science Citation Index) 1900-2019. Search performed 04.04.2019.

Search	Terms	Results
S12	S11 AND SS2 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	6
S11	S7 AND SS10 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	341
S10	S6 AND S4 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	357
S9	S8 AND S6 AND S4 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	0
S8	S5 AND S2 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	503
S7	TS=(<i>*abroad OR sojourn*AND listening OR read* OR writ* OR oral OR spe* OR communi*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	14,511,399
S6	TS=(<i>“language develop*” OR SLA OR “second language acquisition” OR “L2 develop” OR linguistic OR profici*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	146,041
S5	TI=(<i>“language develop*” OR SLA OR “second language acquisition” OR “L2 develop* OR linguistic OR profici*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	40,051
S4	TS=(<i>“study abroad”* OR “year abroad” OR ERASMUS OR “residenc* abroad”</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	6,299
S3	TI=(<i>“study abroad”* OR “year abroad” OR ERASMUS OR “residenc* abroad”</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	3,773
S2	TI=(<i>random* control* trial* OR rct* OR trial* OR review* OR intervent* OR quasi* experimental* OR meta analys*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	1,178,237
S1	TS=(<i>sojourn* OR student* OR “language learner*” OR university*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	1,133,720

Study records

Data management

Any relevant literature found will be uploaded to EPPI reviewer version 4 (found at <http://eppi.ioe.ac.uk/eppireviewer4/>). This software is an internet-based programme which allows for the collection and coding of all material. All review members will have access to the EPPI database. For screening at stages 1 and 2, the coding inclusion and exclusion criteria will be used to screen out any irrelevant study.

Selection process

At stage 1 of the screening process, the review authors will screen on title and abstract against the inclusion criteria. The study's paper will then be found if it is included or any issues arise surrounding its eligibility. In instances of doubt regarding eligibility, a third member of the review team will be asked to make the final judgement, with these disagreements being resolved through discussion. Any reasons for exclusion will be noted. A sample will be done by a second reviewer at each stage.

Data-collection process

Data from studies will be extracted at a number of levels, including demographic information; study design; methodology; outcomes, and results. Two reviewers will independently extract data, and in instances of uncertainty, a third reviewer will again make the judgement call, with all disagreements being resolved through discussion.

Data Items

The variables extracted will include publication status; sample size; study design; participant demographics (e.g., age; gender etc.); length of stay abroad; accommodation type; host country; form of linguistic measure (e.g., oral/written composition; grammar test); outcome skill (speaking, reading etc.); Any variables accounting for such change; results; reasons for exclusion.

Outcomes and prioritisation

The primary outcome is to establish a difference in outcomes or change scores in terms of linguistic ability in language learners. These differences in outcomes or change scores can be measured either through pre/post-tests and/or comparing them with a control group (see study design). The outcomes include anything related to linguistic ability. Outcomes can include but not limited too; written compositions, any form of cloze-test; grammatical testing; vocabulary size; oral composition; quality of pronunciation; written/oral fluency and accuracy; reading scores (mean scores); listening ability (mean scores). Outcomes will cover both the four skills and the lower order language aspects such as grammar, vocabulary and metalinguistic awareness of which can fall under more than one skill. Studies will then be grouped according to the outcome skill they relate to (e.g., speaking; writing, etc.) although it is highly likely that studies will make reference to multiple skills.

The main issue surrounding outcomes is likely to be the use of different instrumentation to measure the same outcome. There are multiple ways to measure, for example, oral proficiency, and as such, this may limit the cohesiveness of results and increases the threat of external validity.

Risk of bias

In order to aid the assessment of risk of bias a modified version of TREND (Transparent Reporting of Evaluations with Non-randomized Designs; found at https://www.cdc.gov/trendstatement/pdf/trendstatement_TREND_Checklist.pdf) will be used for non-randomised studies and correlational studies. For randomised studies, a modified version of CONSORT will be used (found at <http://www.consort-statement.org>).

This checklist includes reviewer judgment of: randomisation type, allocation concealment and procedure, blinding and techniques used to minimise bias in non-randomised studies.

Data synthesis

A narrative synthesis will be undertaken. The structuring of this synthesis will be aided and guided by the *Synthesis Without Meta-analysis (SWiM)* guidelines (Campbell et al., 2020). The aim of the guidelines “is to guide clear reporting in reviews of interventions in which alternative synthesis methods to meta-analysis of effect estimates are used” (Campbell et al., 2020, p. 1). The guidelines, all nine in total, have been designed through collaboration with experienced reviewers, and build on predominantly guidelines set in Preferred Reporting Items for Systematic Reviews and Meta-Analyses and Realist And MEta-narrative Evidence Syntheses: Evolving Standards

Meta-bias(es)

In order to ascertain whether any reporting bias is present in the studies included, we will determine if any such protocol had been made available before participants were found.

Inclusion and Exclusion Criteria

Included	Excluded
Topic: Study abroad, including affiliated organisations such as the British Council.	Topic: Non-study abroad related interventions
Date: No time bias	Date: -
Publication status: All published and unpublished material which is in the public domain	Publication status: -
Study design	Study design
Non-ERASMUS: Any study design where there is a control or comparison group – RCT (individual and cluster); quasi-experiment (interrupted/control time series designs, control group post-test only, control group pre/post-test).	Non-ERASMUS: Case-study designs; designs with only post-test and no control group; basic time series designs. Review articles and non-empirical literature.
ERASMUS: All of the above AND pre-experimental designs (e.g., pre/post-test with no control group)	ERASMUS: Case-study designs; designs with only post-test and no control group. Review articles and non-empirical literature.
Participants: Any undergraduate/postgraduate student undertaking a study abroad as part of their academic degree studies. Control students must be a comparable group (e.g., matched comparisons at baseline) and hold characteristic similarities to those who go abroad at baseline.	Participants: Non-academic learners or are under the age of 18.
Intervention: Studies which include a study abroad of which is longer than five weeks in length.	Intervention: Does not have a study abroad component. A length of stay less than five weeks. Summer school programmes.
Outcomes: Studies which learners are measured at post-test on any linguistic skill and their relevant skill outcome, e.g., speaking, writing, reading, listening, pragmatics. This can be measured through multitude instruments – for example, length of utterances, length of prose, speech/written accuracy/fluency, reading score, listening score, grammatical score. Outcome must be objective (i.e., not self-report)	Outcomes: Measures not looking at linguistic gain e.g., intercultural competency. Outcomes which are self-rated/perceived change e.g., on a scale of 1 – 10 how, much do you believe you have improved.

Appendix B: Systematic Review Search Strings

Web of Science

Search	Terms	Results
S12	S11 AND S2 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	6
S11	S7 AND S10 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	341
S10	S6 AND S4 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	357
S9	S8 AND S6 AND S4 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	0
S8	S5 AND S2 <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	503
S7	TS=(<i>*abroad OR sojourn*AND listening OR read* OR writ* OR oral OR spe* OR communi*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	14,511,399
S6	TS=(<i>“language develop*” OR SLA OR “second language acquisition” OR “L2 develop” OR linguistic OR profici*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	146,041
S5	TI=(<i>“language develop*” OR SLA OR “second language acquisition” OR “L2 develop” OR linguistic OR profici*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	40,051
S4	TS=(<i>“study abroad”* OR “year abroad” OR ERASMUS OR “residenc* abroad”</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	6,299
S3	TI=(<i>“study abroad”* OR “year abroad” OR ERASMUS OR “residenc* abroad”</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	3,773
S2	TI=(<i>random* control* trial* OR rct* OR trial* OR review* OR intervent* OR quasi* experimental* OR meta analys*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	1,178,237
S1	TS=(<i>sojourn* OR student* OR “language learner*” OR university*</i>) <i>Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-SHH, ESCI Timespan=All years</i>	1,133,720

First Search (Article First/ECO)

Search	Terms	Results
S10	S9 AND S7 <i>Databases = Article First, ECO Timespan=All years</i>	51
S9	S6 AND S4 <i>Databases = Article First, ECO Timespan=All years</i>	95
S8	S5 AND S3 AND S2 <i>Databases = Article First, ECO Timespan=All years</i>	0
S7	kw: abroad OR kw: sojourn* AND kw: listening OR kw: read* OR kw: writ* OR kw: oral OR kw: spee* OR kw: communi* <i>Databases = Article First, ECO Timespan=All years</i>	2,389,242
S6	kw: language w develop* OR kw: SLA OR kw: second w language w acquisition OR kw: L2 w develop* OR kw: linguistic OR kw: profici* <i>Databases = Article First, ECO Timespan=All years</i>	51,379
S5	ti: language w develop* OR ti: SLA OR ti: second w language w acquisition OR ti: L2 w develop* OR ti: linguistic OR ti: profici* <i>Databases = Article First, ECO Timespan=All years</i>	19,315
S4	kw: study w abroad OR kw: year w abroad OR kw: ERASMUS OR kw: residenc* w abroad <i>Databases = Article First, ECO Timespan=All years</i>	2,702
S3	ti: study w abroad OR ti: year w abroad OR ti: ERASMUS OR ti: residenc* w abroad <i>Databases = Article First, ECO Timespan=All years</i>	1,709
S2	((ti: random* and ti: control* and ti: trial*) OR ti: rct*) OR ti: trial* OR ti: review* OR ti: intervent* OR (ti: quasi* and ti: experimental*) OR (ti: meta and ti: analys*) <i>Databases = Article First, ECO Timespan=All years</i>	337,143
S1	kw: sojourn* OR kw: student* OR kw: language w learner* OR kw: university* <i>Databases = Article First, ECO Timespan=All years</i>	671,414

Article First 20; ECO 75

EBSCO (British Education Index; ERIC; PsycARTICLES; PsycINFO)

Search	Terms	Results
S9	S6 AND S4 <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	1,173
S8	S7 AND S5 AND S3 AND S2 AND S1 <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	32
S7	TX=(abroad OR sojourn*AND listen* OR read* OR writ* OR oral OR spee* OR communi*) <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	2,087,579
S6	TX=(“language develop* OR SLA OR “second language acquisition” OR “L2 develop* OR profici* OR linguistic) <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	236,414
S5	TI=(“language develop* OR SLA OR “second language acquisition” OR “L2 develop* OR profici* OR linguistic) <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	28,086
S4	TX=(“study abroad” OR “year abroad” OR ERASMUS OR “residenc* abroad”) <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	16,631
S3	TI=(“study abroad”* OR “year abroad” OR ERASMUS OR “residenc* abroad”) <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	1,705
S2	TI=(random* control* trial*) OR rct*) OR trial* OR review* OR intervent* OR quasi* experimen* OR (meta analys*) <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	303,125
S1	TX=(sojourn* OR student* OR “language learner*” OR university*) <i>Databases = British Education Index; ERIC; PsycARTICLES; PsycINFO, Timespan=1900-2019</i>	3,698,889

British Education Index: 140; ERIC: 603; PsycArticles: 145; PsycInfo: 285

ProQuest (Dissertation & Theses Global AND Social Science Premium Collection)

Search	Terms	Results
S8	S7 AND S5 <i>Databases=ProQuest Dissertations & Theses Global, Social Science Premium Collection, Timespan = 1900-2019</i>	474
S7	S4 AND S3 <i>Databases=ProQuest Dissertations & Theses Global, Social Science Premium Collection, Timespan = 1900-2019</i>	475
S6	S5 AND S4 AND S3 AND S2 AND S1 <i>Databases=ProQuest Dissertations & Theses Global, Social Science Premium Collection, Timespan = 1900-2019</i>	5
S5	su(abroad) OR su(sojourn*) AND su(listening) OR su(read*) OR su(writ*) OR su(oral) OR su(spec*) OR su(communi*) <i>Databases=ProQuest Dissertations & Theses Global, Social Science Premium Collection, Timespan = 1900-2019</i>	1,315,255
S4	su(“language develop*”) OR su(SLA) OR su(“second language acquisition”) OR su(“L2 develop*”) OR su(profici*) OR su(linguistic) <i>Databases=ProQuest Dissertations & Theses Global, Social Science Premium Collection, Timespan = 1900-2019</i>	74,898
S3	su(“study abroad”) OR su(“year abroad”) OR su(ERASMUS) OR su(“residenc* abroad”) <i>Databases=ProQuest Dissertations & Theses Global, Social Science Premium Collection, Timespan = 1900-2019</i>	20,088
S2	su(random* control* trial*) OR su(rct*) OR su(trial*) OR su(review*) OR su(intervent*) OR su(quasi*) OR su(experimental*) OR su(meta analys*) <i>Databases=ProQuest Dissertations & Theses Global, Social Science Premium Collection, Timespan = 1900-2019</i>	1,311,875
S1	su(sojourn*) OR su(student*) OR su(“language learner*”) OR su(university) <i>Databases=ProQuest Dissertations & Theses Global, Social Science Premium Collection, Timespan = 1900-2019</i>	2,505,478

Scopus

Search	Terms	Results
S8	S4 AND S3	383
S7	S5 AND S4 AND S3 AND S2	0
S6	S5 AND S4 AND S3 AND S2 AND S1	0
S5	Title, Abstract, Keywords = (abroad OR sojourn*AND listen* OR read* OR writ* OR oral OR spec* OR communi*)	10,272
S4	Title, Abstract, Keywords = (language develop* OR SLA OR “second language acquisition” OR L2 develop* OR linguistic OR profici*)	204,417
S3	Title, Abstract, Keywords = (“study abroad*” OR “year abroad” OR ERASMUS OR “residenc* abroad”)	6,223
S2	Title, Abstract, Keywords = (random* control* trial* OR rct* OR trial* OR review* OR intervent* OR “quasi experimen*”) OR “meta analys*”)	837,863
S1	Title, Abstract, Keywords = (sojourn* OR student* OR “language learner*”) OR university*)	2,678,860

Appendix C: Data Extraction Tool

Publication reference	Hessel (2016)
Description of intervention (accommodation, length of intervention, mode of learning e.g., studying)	All learners were completing an ERASMUS study abroad programme. They came from 44 institutions across Germany No information given on the accommodation. Participants had on average 11.3 academic contact hours per week
Description of design (QED, pre/post-test) If a QED, how were the two groups formed (e.g., assigned grouping or self-selection)	QED implemented with two intervention groups (a long-stay/short-stay group) and a stay-at-home comparison group. Students were recruited via an email sent to the ERASMUS office of institutions – students self-selected themselves to take part.
Description of participants in intervention group (e.g., learning experience?) How many participants in group?	Group 1: 54 ERASMUS students who studied abroad for one year. Group 2: 45 ERASMUS students who studied abroad for one term. All students had extensive background of English learning (8-9 years) Had two terms of English language instruction at university Rated upper-intermediate to advanced on most recent IELTS/TOEFL/CAE result
Description of participants in comparison group (e.g., learning experience?) How many participants in group?	Stay-at-home control group ($n = 44$) and who were either unsuccessful or had withdrawn their ERASMUS application. At baseline, the comparison group were equivalent on all tested variables apart from 'prior extended period abroad'. Significantly more individuals in the comparison group had experienced an extended stay of more than 1-month in an English-speaking country.
Instruments/Linguistic component (e.g., oral, writing, lexical richness, syntactic complexity) Frequency of datapoints	Language instrument comprised of 5 C-tests which measured global proficiency. Instruments were disseminated three times across the study abroad period. Pre-sojourn, mid-sojourn, post-sojourn.
Results as reported – please provide quotations + page numbers	During the first 3 months of the study abroad, ERASMUS students made significantly higher gains than the at-home control group, with the learning context being found to be a significant predictor of language change after controlling for baseline characteristics ($p = .001$). When comparing the long-term group and control group for the next six months, the differences in gain scores between the two groups were not significantly different indicating a slow-down in growth for the ERASMUS group.
Conclusions as reported	Sojourning can be said to account for language change in the first three months of the study abroad. Extended stays abroad appear less beneficial and it can be said that sojourning does not result in significant growth after the first semester abroad.

Quality assessment

Are the conclusions a fair reflection of the results analysed?	Yes
How have the intervention and control groups been formed? Is sampling bias an issue?	Self-selection – there may be a concern that more proficient learners show greater interest in the study.
Has attrition been explored and how this may impact overall results?	Attrition has not been explored
Have the authors pre-specified which outcome serves as the primary outcome?	No primary outcomes shown
Other	

Publication reference	Hessel & Vanderplank, 2018
Description of intervention (accommodation, length of intervention, mode of learning e.g., studying)	All learners were completing an ERASMUS study abroad programme. They came from 44 institutions across Germany No information given on the accommodation. Participants had on average 11.3 academic contact hours per week
Description of design (QED, pre/post-test) If a QED, how were the two groups formed (e.g., assigned grouping or self-selection)	QED implemented with two intervention groups (a long-stay/short-stay group) and a stay-at-home comparison group. Students were recruited via an email sent to the ERASMUS office of institutions – students self-selected themselves to take part.
Description of participants in intervention group (e.g., learning experience?) How many participants in group?	96 successful ERASMUS applicants 52 were studying the UK for the academic year (group 1) 44 were studying abroad for 1 semester (group 2) Had extensive learning histories of English (an average of 4 terms) English proficiency upper intermediate to advance.
Description of participants in comparison group (e.g., learning experience?) How many participants in group?	40 students who had been unsuccessful in gaining an ERASMUS or had withdrawn their application (group 3)
Instruments/Linguistic component (e.g., oral, writing, lexical richness, syntactic complexity) Frequency of datapoints	Test general proficiency through a c-test – consisted of 5 short texts on random topics. Completed online only and was measured over three timepoints: T1 (September); T2 (December) and T3 (June)
Results as reported – please provide quotations + page numbers	No significant differences at baseline between the three groups “The paired samples t-tests show highly significant mean gains in L2 proficiency among students in both abroad groups (group 1 <i>mean</i> =8.86, <i>SD</i> =6.56, min.: -8, max.: 22; group 2 <i>mean gain</i> =8.23, <i>SD</i> =9.89, min.: -11, max.: 37; both <i>p</i> <.001). The mean gain within groups 1 and 2 was very similar and medium in size (Cohen, 1988). However, the greater variability in the mean proficiency gain within abroad group 2 suggests that individual differences in linguistic progress were more pronounced among students who expected to stay abroad for one term only. No significant changes in overall English proficiency were observed among students in group 3 who continued to study in the home country.” p.204
Conclusions as reported	The current study has provided novel evidence on overall L2 proficiency gain as an outcome of participation in study abroad programmes of one term or longer. P. 212 The results therefore substantiate the hypothesis that studying abroad for one term brings significant linguistic benefits for advanced L2 learners with regard to developing overall L2 proficiency. At the same time, the results put a question mark over the added benefit of a longer stay and highlight the need for institutions to facilitate sustained linguistic progress among their international students. P.212

Quality assessment

Are the conclusions a fair reflection of the results analysed?	Yes
How have the intervention and control groups been formed? Is sampling bias an issue?	The participants self-selected themselves to take part in the study. The control group were made up of individuals who wanted to do ERASMUS but were unsuccessful. Demographic results showed no significant differences in characteristics although those in the control group had more prior experience of being abroad.
Has attrition been explored and how this may impact overall results?	Attrition not noted – assumed to use listwise deletion
Have the authors pre-specified which outcome serves as the primary outcome?	No pre-defined outcomes
Other	

Publication reference	Jochum, 2014
Description of intervention (accommodation, length of intervention, mode of learning e.g., studying)	1 semester long (3 months) all SA students went to Peru and lived with host families. SA students were studying and took four to five language classes at a university.
Description of design (QED, pre/post-test) If a QED, how were the two groups formed (e.g., assigned grouping or self-selection)	Pre/post-test, quasi-experimental design. No random assignment between the control and experimental groups.
Description of participants in intervention group (e.g., learning experience?) How many participants in group?	9 students in SA group Purposeful criterion sampling used – Spanish minor/major; NS English; No previous study abroad exp. And enrolment in a Spanish class at university level
Description of participants in comparison group (e.g., learning experience?) How many participants in group?	9 students in AH group Enrolled in one or more Spanish classes on a US campus
Instruments/Linguistic component (e.g., oral, writing, lexical richness, syntactic complexity) Frequency of datapoints	Oral proficiency measured by the ACTFL Oral Proficiency Interview Pre-test: Completed by both groups – administered within six weeks of the SA students’ departure Post-test: Completed by both groups – administered within six weeks of the end of the spring semester. Taken under exam conditions
Results as reported – please provide quotations + page numbers	No significant difference at baseline in oral proficiency ($F(1,16) = .566, p = .463$). A one-way ANOVA was conducted to determine if the post-test mean proficiency scores differed between the two groups; results revealed significance ($F(1,16) = 7200, p = .016$). It is also important to note that a series of paired-sample t tests indicated that the pre/post-test improvement was statistically significant within the SA group ($t(8) = -3.59, p = .007$) but inconclusive within the AH group ($t(8) = -2.29, p = .051$). p. 100
Conclusions as reported	In fact, these findings suggest that students who study abroad for a semester can not only increase their levels of oral proficiency but can do so at a rate that is significantly higher than studying on campus. Overall, this study confidently supports the importance of going abroad in order to achieve higher levels of proficiency. P. 101

Quality assessment

Are the conclusions a fair reflection of the results analysed?	Yes
How have the intervention and control groups been formed? Is sampling bias an issue?	Students self-selected to take part in the study – participants were purposefully selected. Students equivalent at baseline.
Has attrition been explored and how this may impact overall results?	No attrition reported
Have the authors pre-specified which outcome serves as the primary outcome?	Only one outcome measured – oral proficiency
Other	

Publication reference	Li, 2014
Description of intervention (accommodation, length of intervention, mode of learning e.g., studying)	No description of intervention is given. Only that sojourners went abroad for 8 weeks and that the programme in China was equivalent to the regular course offered by the US university.
Description of design (QED, pre/post-test) If a QED, how were the two groups formed (e.g., assigned grouping or self-selection)	Overall proficiency: pre-test/post-test – beginners did not have a pre-test as they had no previous experience. Pre-test scores were formed from the scores achieved at the end of the semester prior to going abroad. Reading ability: pre/post-test. SA pre-test taken one week before departure and post-test taken during the first week of the fall semester (end of 8-week summer program) AH pre-test in first week of fall semester and post-test in last week of fall semester.
Description of participants in intervention group (e.g., learning experience?) How many participants in group?	SA group split into three proficiencies – 9 were in beginners; 15 intermediate; 11 advanced SA group spent 8 weeks abroad (1 semester) Native speakers of English
Description of participants in comparison group (e.g., learning experience?) How many participants in group?	AH group split into three proficiencies – 15 were in beginners; 13 intermediate; 10 advanced Native speakers of English Those who had previous experience of SA programmes in China were excluded from AH. Heritage students of Chinese were excluded from both groups
Instruments/Linguistic component (e.g., oral, writing, lexical richness, syntactic complexity) Frequency of datapoints	Chinese language proficiency test. This was measured out of 100 and covered listening, grammar, reading, translation and writing. Reading comprehension test – two passages with five multiple choice questions for each. Highest score was 10. Think aloud task Observation Reading strategies questionnaire Semi-constructive retrospective interview Language Contact Profile and Reading Language Contact Profile
Results as reported – please provide quotations + page numbers	Overall proficiency: There was no significant difference on the scores between the SA and AH groups among beginners, $F(1, 24) = 2.342, p = .101$. The scores of the SA groups were significantly higher than the AH groups at the intermediate and advanced level respectively, $F(1, 27) = 31.634, p < .001$; $F(1, 20) = 12.804, p = .003$. p. 81 Reading: There was no significant difference on the scores between the two learning contexts for the beginners, $F(1, 24) = 64.238, p = 42.093$. The scores of the SA groups were significantly higher than the AH groups among the intermediate students, $F(1, 27) = 41.382, p < .001$. There was no significant difference among the advanced learners, $F(1, 20) = 12.804, p = 2.103$. P.82
Conclusions as reported	To the first research question, we found that overall, the language proficiency of the SA groups was significantly higher than the AH groups across the three different proficiency levels. Closer examination revealed that there was no significant difference among the beginners. However, the advantage of the SA context over the AH context was significant among the intermediate and advanced students. P. 86 Appears to be a threshold required to make learning gain – beginners did make much gain – learners require a certain level of knowledge.

Quality assessment

Are the conclusions a fair reflection of the results analysed?	Yes – authors have not over reported
How have the intervention and control groups been formed? Is sampling bias an issue?	No information given on how the groups have been formed.
Has attrition been explored and how this may impact overall results?	No information on attrition has been given
Have the authors pre-specified which outcome serves as the primary outcome?	No, the authors have treated each variable in its original order.
Other	

Publication reference	Llanes & Munoz (2013)
Description of intervention (accommodation, length of intervention, mode of learning e.g., studying)	ERASMUS programme with a length of stay of 2-3 months – participants stayed in the UK and Ireland Amount of class contact varied greatly and was determined by the number of credits the students were registered on at the home university. Students reported various living arrangements, with/without L2 speakers.
Description of design (QED, pre/post-test) If a QED, how were the two groups formed (e.g., assigned grouping or self-selection)	Pre/post-test QED. 25 of the 46 SA students were collected at their host university on their first day of the semester. Post-test data was collected 2 months later. The remaining 21 had their pre-test data collected at the home university, one week before departure. Post-test data was collected a week after returning home.
Description of participants in intervention group (e.g., learning experience?) How many participants in group?	46 adult learners of English on an ERASMUS programme Had a mean age of 20.9 and had an average length of previous exposure of 1,620 hours 92% of individuals were Spanish/Catalan bilinguals
Description of participants in comparison group (e.g., learning experience?) How many participants in group?	20 adult learners of English remaining at home. It was an intact class of English majors at a Catalan university. Had a mean age of 20.9 and had an average length of previous exposure of 1,620 hours
Instruments/Linguistic component (e.g., oral, writing, lexical richness, syntactic complexity) Frequency of datapoints	Written data: asked to write a piece of writing more than 7 lines long in 15 mins. Topic was on “my life, past, present, and future expectations”. Oral data: Interviewed in English through a semi structured interview. This served as a warm-up task. Main task was a picture-elicited narrative task in which participants shown six pictures. Individuals given 1 minute to examine and formulate a story prior to retelling and recording. Language Contact Profile also used to measure contact – only 21 of the 46 adults returned this. Writing measurements: Writing fluency (words per t-unit); lexical complexity (lexical richness); syntactic complexity clauses per t-unit); and accuracy (errors per t-unit) Oral measurements: Same as written measures but oral fluency measured via pruned syllables per minute.
Results as reported – please provide quotations + page numbers	For the adults, paired samples <i>t</i> tests also revealed that the group of SA adults scored significantly higher in one out of the four oral variables examined, namely fluency, $t(45) = -7.507, p = .000$, but in none of the written variables. By comparison, the AH group of adults did not score significantly higher in any of the oral measures, but they did in written lexical complexity, $t(17) = -3.383, p = .004$. p. 74 It was observed in the follow-up tests that SA participants scored higher than those who remained at home in all measures analysed, as the marginal means were higher for the group of SA participants than for the group of AH participants. P.76 Improvement in oral fluency and lexical complexity was quite uniform across the SA group. P.76 To sum up, the MANCOVA results for learning context indicate that the SA context was more beneficial than the AH context for the improvement of oral skills, but not as much for improving writing skills, as measured in this study. P.76 SA adults had the highest gains in oral lexical complexity, and AH adults had the highest gains in written fluency and syntactic complexity. Thus the SA setting seems to be more beneficial for children in terms of the improvement of oral skills, whereas the AH context seems to foster the development of writing skills, especially for adults. P.79

Conclusions as reported	The comparative analyses indicate that the SA context is more beneficial than the AH context, particularly for the improvement of participants' L2 oral skills. P.83 SA appears more beneficial for oral skills rather than writing.
-------------------------	---

Quality assessment

Are the conclusions a fair reflection of the results analysed?	Yes
How have the intervention and control groups been formed? Is sampling bias an issue?	No discussion on how the groups were formed
Has attrition been explored and how this may impact overall results?	Attrition not specifically mentioned
Have the authors pre-specified which outcome serves as the primary outcome?	No, oral change always described before writing gain
Other	

Publication reference	Segalowitz & Freed (2004)
Description of intervention (accommodation, length of intervention, mode of learning e.g., studying)	Individuals were studying abroad and the study lasted 13 weeks (1 semester)
Description of design (QED, pre/post-test) If a QED, how were the two groups formed (e.g., assigned grouping or self-selection)	Pre/post-test, QED with intervention and comparison group. Both groups took pre and post-test. No info on how the groups were formed although there was a set criteria on who could partake in the study.
Description of participants in intervention group (e.g., learning experience?) How many participants in group?	22 students went abroad. They were living in Spain. They were studying three language class per week. Mean age = 20.68; 18 females and 4 males. 13 students took an additional one or two complementary Spanish classes per week on Spanish society and culture.
Description of participants in comparison group (e.g., learning experience?) How many participants in group?	18 students stayed AH. They were studying Spanish at a university in Colorado. They were studying one language class per week. Mean age = 23.39; 14 females and 4 males.
Instruments/Linguistic component (e.g., oral, writing, lexical richness, syntactic complexity) Frequency of datapoints	Language contact: Pre-test version of the Language Contact Profile measuring contact with L2 Oral performance: Measured via an Oral Proficiency Interview capturing the following: total number of words, duration, longest turn, speech rate, mean length of run without silent pauses of 400ms or longer, mean length of run without filled pauses, and longest fluent run. Cognitive performance: Lexical access – timed, two-alternative forced choice judgement tests and attention control – timed responses to indicate which of the three words on a screen matched or did not match a sample stimulus. Two datapoints – gap of 13 weeks. Both groups complete the same measures.
Results as reported – please provide quotations + page numbers	As can be seen from Table 1, only the students in the SA context made significant gains in oral performance, and they did so on four of the seven oral measures: Turn, Rate, Filler-free, and Fluent-run. The measures Turn, Rate, Filler-free, and Fluent-run were consequently submitted to two-way mixed ANOVAs with the between factor being context AH, SA and the within factor being time pre-test, post-test. Significant interaction effects indicating that the SA group changed significantly more than the AH group were found for three of the variables: Turn; Rate; and Filler-free. The interaction effect was marginally significant and weak for Fluent-run; These results suggest that the SA group made greater oral gains than the AH group. Thus, students in the SA context made gains on five of the eight oral measures—OPI, Turn, Rate, Filler-free, and Fluent-run—and students in the AH context did not. Variation in the SA group’s gains did not appear to reflect global contact time with the language, either in or out of class. No context effects for any cognitive measure as shown through no significant interaction effects for abroad vs. home
Conclusions as reported	The results indicated that, compared to the AH context, learning in the SA context led to significantly greater oral performance gains. This was seen with respect to pretest-posttest differences on two general oral performance variables—OPI and longest speaking turn—and on three oral fluency measures—speech rate, mean length of speech run not containing filled

	<p>pauses, and longest fluent run not containing silent hesitations or filled pauses, all indicating greater gains for the SA students. P. 193</p> <p>Authors do present alternative reasoning for change: SA students had more language courses per week (3 vs 1)</p> <p>Thus, in answer to the second set of questions, these results indicate that, overall, the two learning contexts led to similar gains in fluency-relevant cognitive processing abilities and that the relationship between these gains and time-on-task variables was complex. P. 194</p>
--	--

Quality assessment

Are the conclusions a fair reflection of the results analysed?	Yes
How have the intervention and control groups been formed? Is sampling bias an issue?	No information provided on how the groups were formed.
Has attrition been explored and how this may impact overall results?	<p>7 individuals were dropped from analysis.</p> <p>“Criteria for retaining students in the study were the following: English had to be their L1; they had to have studied Spanish as an L2 for at least two semesters; they had to have never studied Spanish abroad before; Spanish was not their heritage language; no one spoke Spanish in their home; and they had to complete all interviews and tests described in the following section.” P. 178</p> <p>Cognitive tests had for some a smaller sample size. Any individual with an error rate of 21% or more were dropped.</p>
Have the authors pre-specified which outcome serves as the primary outcome?	No order specified. Oral proficiency was always explored before cognitive ability.
Other	

Publication reference	Serrano, Llanes, Tragant (2011)
<p>Description of intervention (accommodation, length of intervention, mode of learning e.g., studying)</p>	<p>SA: While in the UK, the majority of the students in this study (76%) had a total of 8-12 h a week of classes in English (including English language classes). Most of the SA participants stayed in houses with other students (60%), while 20% stayed in halls of residence and another 20% with families. I think this is linked to Llanes & Serrano (2011)</p> <p>Intensive: the programme offered 110h of instruction in 5-h sessions over four and half weeks in the summer.</p> <p>Semi-intensive course: same number of hours (110) but over 3 months – average of 10 hours of instruction in 2.5-h sessions.</p>
<p>Description of design (QED, pre/post-test)</p> <p>If a QED, how were the two groups formed (e.g., assigned grouping or self-selection)</p>	<p>QED, pre/post-test – no information given on how the groups were formed, assume self-selection.</p>
<p>Description of participants in intervention group (e.g., learning experience?)</p> <p>How many participants in group?</p>	<p>Group 1: 25 students went abroad all to the same university. L1 was Spanish/Catalan. All had gained a scholarship to study abroad through ERASMUS.</p>
<p>Description of participants in comparison group (e.g., learning experience?)</p> <p>How many participants in group?</p>	<p>Group 2: intensive programme – ranged from intermediate to advance</p> <p>Group 3: semi-intensive – all intermediate proficiency</p> <p>All students were aged between 18-23. All the students are comparable in terms of motivation and previous experience with English, according to a background questionnaire they completed. P. 136</p>
<p>Instruments/Linguistic component (e.g., oral, writing, lexical richness, syntactic complexity)</p> <p>Frequency of datapoints</p>	<p>Written data: asked to write a piece of writing more than 7 lines long in 15 mins. Pre-test topic was “my best friend”</p> <p>Main task was a picture-elicited narrative task Because of practical reasons, the oral task was performed by a subgroup of students chosen randomly (N = 12 in the semi-intensive course; N = 43 in the intensive programme).</p> <p>Writing measurements: Writing fluency (words per t-unit); lexical complexity (lexical richness); syntactic complexity clauses per t-unit); and accuracy (errors per t-unit)</p> <p>Oral measurements: Same as written measures but oral fluency measured via pruned syllables per minute.</p> <p>Measurements between pre and post-test for Gr 2 and Gr 3 was 80 hours.</p> <p>Measurements for Gr 1 and Gr 3 was around 70 days.</p> <p>In order to compare the SA students with the AH intensive learners, the former per- formed the post-test 15 days after the pre-test, which was the same lapse of time between the administration of pre-test and post-test for the AH intensive group. To facilitate comparison with the AH semi-intensive programme, the SA learners wrote another composition on a similar topic (“My best friend in Southampton”) and told the oral narrative again approximately two months after the pre-test, which was the time between both tests for the AH semi-intensive group</p> <p>At post-test: Written data: asked to write a piece of writing more than 7 lines long in 15 mins. Post-test topic was “someone I admire”</p>

<p>Results as reported – please provide quotations + page numbers</p>	<p>AH Intensive and SA: The descriptive statistics show that the scores obtained by the learners in the SA context in the post-test were, with the exception of lexical complexity, slightly higher than those obtained by the learners in the AH intensive program. Nevertheless, the results of the MANCOVA, after controlling for pre-test performance, indicate that no differences existed between the learners in AH intensive programme (N =69) and in the SA context (N = 24) on the combined dependent variables: $F(4, 84) = 1.05, p = .388, \text{Wilks' Lambda} = .952, \text{partial eta squared} = .048$.</p> <p>The results of the MANCOVA were also similar to those of the written production task, in that no significant differences existed between the two contexts on the combined dependent variables: $F(4, 58) = .196, p = .940, \text{Wilks' Lambda} = .987, \text{partial eta squared} = .013$. p. 139</p> <p>AH Semi-Intensive and SA: The results of the MANCOVA indicate that there was a statistically significant difference between the students abroad and those in the semi-intensive course on the combined dependent variables: $F(4, 53) = 7.64, p < .001, \text{Wilks' Lambda} = .634, \text{partial eta squared} = .366$. Considering the results of the dependent variables separately, the variables in which significantly higher scores were obtained in the SA context were W/T [$F(1, 60) = 4.12, p = .047, \text{partial eta squared} = .069$] and Guiraud's Index [$F(1, 60) = 19.62, p < .001, \text{partial eta squared} = .260$]. The students' performance in terms of syntactic complexity and accuracy was not significantly different for the two contexts.</p> <p>According to the outcome of the MANCOVA analysis, there was also a statistically significant difference favouring the students abroad on the combined dependent variables: $F(4, 28) = 3.28, p = .025, \text{Wilks' Lambda} = .682, \text{partial eta squared} = .318$. Examining the variables separately, there were significant differences in terms of fluency (Syll/ min): $F(1, 35) = 4.92, p = .034, \text{partial eta squared} = .138$; and lexical complexity (Guiraud's Index): $F(1, 35) = 4.32, p = .046, \text{partial eta squared} = .122$. As was the case for the written production task, learners' oral syntactic complexity and accuracy after two months abroad or after receiving two months of instruction at home were comparable. P. 140</p>
<p>Conclusions as reported</p>	<p>After two months abroad, the learners in the present study demonstrated a more advanced performance in terms of some variables of written and oral production than their peers spending the same period of time in a semi-intensive course AH. In contrast, the students' L2 written and oral production after spending 15 days abroad or the same period in an intensive course at home was similar. P. 140</p> <p>In contrast with the learners in the AH semi-intensive context, the learners in the AH intensive programme do not appear to be at a disadvantageous position with respect to their peers abroad. After controlling for pre-test scores, there were no differences in the measures of written and oral production under study between the learners following an intensive course AH and the learners abroad. P.140</p>

Quality Assessment

<p>Are the conclusions a fair reflection of the results analysed?</p>	<p>Yes</p>
<p>How have the intervention and control groups been formed? Is sampling bias an issue?</p>	<p>No information given on how the groups were formed – it is likely self-selection on to the study once context criteria had been met.</p>
<p>Has attrition been explored and how this may impact overall results?</p>	<p>1 student dropped from SA context because they couldn't complete the post-test.</p>
<p>Have the authors pre-specified which outcome serves as the primary outcome?</p>	<p>No – outcomes were not specified although oral outcomes described before writing.</p>
<p>Other</p>	

Appendix D: Systematic Review Mapping Table (40 studies)

Author(s) and Title	Study Design	Intervention	Outcomes and Instrument	Participants	Key Findings
Avello, P. (2014). Assessing learners' changes in foreign accent during Study Abroad ¹	Longitudinal, pre, post-test design Own matched pairs over a 15-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English university.	Perceived foreign accent (pronunciation) as measured by a rating task	8 L3 (aged 17-21) learners of English All had similar experience of learning English (roughly 8 years) Previous learning conducted in domestic classroom environment (700-800 hours exposure to L2) All at B2 (Upper intermediate) at the beginning of the intervention.	Increase in FA ratings observed during T1-T2 period. Slight decrease between T2 and T3. This suggests an improvement in the NNSs' degree of FA during the SA period. Change across time was not sig. Foreign accent rating tended to remain similar across time. No evidence of sig. improvement in accent during SA
Avello, P., Mora, J., & Pérez-Vidal, C. (2012). Perception of FA by non-native listeners in a study abroad context ¹	Pre/post-test, no control group design. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English university.	Perceived foreign accent (pronunciation) as measured by a rating task (pre/post SA) and a paired comparison task	23 L3 (aged 17-21) learners of English All had similar experience of learning English (roughly 8 years) Previous learning conducted in domestic classroom environment (700-800 hours exposure to L2) All at B2 (Upper intermediate) at the beginning of the intervention.	Rating task: Non-sig. change although slight decrease in perceived foreign accent Paired-comparisons: post-test samples had been perceived more native like – improvement in FA Accuracy: sig. gains in pronunciation accuracy after SA
Avello, P., & Lara, A. (2014). Phonological development in L2 speech production during study abroad programmes differing in length of stay ¹	Pre/post-test, no AH comparison group. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending either a compulsory 3 months or 6 months studying at an L1 English university.	Pronunciation accuracy measured via acoustic analyses with measures for vowel quality, vowel duration and Voice Onset Time	33 L3 learners of English. Short stay: <i>n</i> = 25; Long stay: <i>n</i> = 8 All had similar experience of learning English (roughly 8 years) Previous learning conducted in domestic classroom environment (700-800 hours exposure to L2) All at B2 (Upper intermediate) at the beginning of the intervention.	Production accuracy of sounds not particularly sensitive to an extended period abroad. No clear improvement towards a more native like performance in both vowel quality and vowel duration 3 or 6 months abroad had little differing impact on improved VOT native like qualities.

Beattie, J., Valls-Ferrer, M., & Pérez-Vidal, C. (2014). Listening performance and onset level in formal instruction and study abroad ¹	Longitudinal, pre, post-test design Own matched pairs over a 30-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English university.	Listening ability Recording was an authentic recording from interview and learners answered three comprehension question types (5 multiple choice; 5 information gap questions; 5 True/False questions)	75 L3 learners of English. All at B2 (Upper intermediate) at the beginning of the intervention. Age range from 17 to 25	Participants made sig.ly greater improvement in their listening skills during SA than at AH
Juan-Garau, M. (2014). Oral accuracy growth after formal instruction and study abroad ¹	Longitudinal, pre, post-test design Own matched pairs over a 30-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English university.	Oral accuracy Measured via a role play, using a two-way, open ended role play where negotiation for meaning was required	43 L3 learners of English. All at B2 (Upper intermediate) at the beginning of the intervention. Average age was 18 at beginning of study	Sig.ly fewer errors post SA in both T-unit and clauses. No improvement in error rate AH Word order errors increased slightly AH and decreased post SA. Agreement errors were also reduced post SA only. Negatives improved in both contexts
Juan-Garau, M., & Pérez-Vidal, C. (2007). The effect of context and contact on oral performance in students who go on a stay abroad ¹	Longitudinal, pre, post-test design Own matched pairs over a 24-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English university.	Oral complexity, accuracy and fluency Oral data collected via a role play task which was a two-way interaction and required negotiation for meaning	12 L3 learners of English.	SA has positive effect on oral ability with one exception: proportion of subordinates which appears to benefit more from AH context No aspect of oral ability sig.ly improved in the AH context. During SA period, learners sig. improved words per clause and words per sentence No other measure showed sig. change
Juan-Garau, M., Salzar-Noguera, J., & Prieto-Arranz, J. (2014). English L2 learners' lexico-grammatical and motivational development at-home and abroad ¹	Longitudinal, pre, post-test design Own matched pairs over a 15-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English university.	Lexico-grammatical competence measured via a cloze test and a sentence rephrasing test	57 L3 learners of English. Ages ranged between 17 and 25	Cloze test: no sig. differences between T1 and T2 but sig. difference at post SA (T3) suggesting sig. progress during SA Sentence rephrasing: Sig. increase at-home and SA. SA change was found to be non-sig.ly different to AH change

Lara, R., Mora, J., & Pérez-Vidal, C. (2015). How long is long enough? L2 English development through study abroad programmes varying in duration ¹	Pre/post-test, no AH comparison group with non-random group assignment. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 or 6 months studying at an L1 English university.	Oral complexity, accuracy and fluency Oral data collected via a role play task which was a two-way interaction and required negotiation for meaning	47 L3 learners of English. Short stay: <i>n</i> = 33; Long stay: <i>n</i> = 14 All at B2 (Upper intermediate) at the beginning of the intervention.	Six months stay no more beneficial than three months Only three-month group showed sig. change in oral fluency and accuracy. No changes in lexical or syntactic complexity for either group over the allotted time
Mora, J. (2014). The role of onset level on L2 perceptual phonological development after formal instruction and study abroad ¹	Longitudinal, pre, post-test design Own matched pairs over a 36-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English university.	Learner perceptual phonological competence assessed by an auditory discrimination task – testing ability to discern difference between vowel and consonant pairs	66 L3 learners of English	Sig. gains in phonological perception only occurred between T1 and T2 -no sig. change between T2 and T3 Those who made most gains between T1 and T2 made least gain between T2 and T3 SA period did not necessarily have a positive impact on speech perception skills
Pérez-Vidal, C. & Juan-Garau, M. (2011). The effect of context and input conditions on oral and written development: A Study Abroad perspective ¹	Longitudinal, pre, post-test design Own matched pairs over a 36-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English speaking university.	Oral and written ability Writing: timed written composition in exam-like conditions Oral: dyadic role play requiring negotiation for meaning	35 L3 learners of English completed writing measure; 20 completed oral measures All at B2 (Upper intermediate) at the beginning of the intervention.	Writing: At post SA, students wrote more fluently and with greater complexity. While at-home (between T1-T2), the participants sig.ly decreased in written fluency and complexity Oral: At post SA, students spoke sig.ly more fluency and with accuracy. Grammatical and lexical complexity increased, albeit non-sig.ly
Pérez-Vidal, C., & Barquin, E. (2014). Comparing progress in academic writing after formal instruction and study abroad ¹	Longitudinal, pre, post-test design Own matched pairs over a 15-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English speaking university.	Writing proficiency as measured via a written composition on the domains of Complexity, Accuracy and Fluency	107 L3 learners of English (73 completed three timepoints; 34 completed all four timepoints) Ages ranged between 17 and 25	At post SA, learners made sig. gains in writing fluency (more words per minute), lexical diversity (fewer repeated words) and writing accuracy (fewer errors per word) than changes made at-home

Trenchs-Parera, M. (2009). Effects of Formal Instruction and a Stay Abroad on the Acquisition of Native-Like Oral Fluency ¹	Longitudinal, pre, post-test design Own matched pairs over a 15-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English speaking university.	Oral fluency measured via seven dysfluency phenomena (i.e., disruptions to the flow of speech)	19 L3 learners of English randomly selected from a group of volunteer students.	T1: sig. NNS behaviour in two dysfluency phenomena compared to NS. T2: Before SA, NNS now sig.ly differ in five measures (i.e., become less native-like). T3: Post SA – slightly corrects this behaviour; NNS now sig.ly differ in four measures. NNS paused less than NS, had more self-repetitions, used more single-word fillers and used more lexicalised phrases
Valls-Ferrer, M., & Mora, J. (2014). L2 fluency development in formal instruction and study abroad ¹	Longitudinal, pre, post-test design Own matched pairs over a 15-month period. Participation voluntary.	ERASMUS L1 Spanish/Catalan learners of English spending a compulsory 3 months studying at an L1 English speaking university.	Oral fluency: semi-guided interview performed in pairs Speech rate, articulation rate, phonation-time ratio, mean length of runs, pause frequency, pause duration	27 L3 learners of English. Mean age at time of study was 18.3 (17-25)	Largest gains found during SA period. Gains seen in SA period were sig.ly larger than FI period on SR, AR & MLoR, PhonRat and PauseDur also improved but this was not sig. PauseFreq showed gains only in FI period with post SA change being greater than pre-test. 3-month provides learner with sig. benefits in oral fluency
Edmonds & Gudmestad. (2018). Gender marking in written L2 French ²	Longitudinal, pre, post-test design Own matched pairs undertaken pre, during and post SA stay. Participation voluntary.	ERASMUS L1 English learners of French spending a compulsory academic year living in a French speaking country	Gender marking (grammar) tested through written task. Corpus coded based on every instance a full noun was modified by determiner or adjective	20 L2 learners of French All had obtained an end-of-secondary-education qualification Average age was 20 (19-21) and had studied on average 10.45 years of French instruction	Linear improvement across time – during stay abroad trend towards more native like competencies compared to at-home. Improvement seen during SA maintained eight months later
Huensch, A., & Tracy-Ventura, N. (2017). Understanding second language fluency behaviour: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time ²	Longitudinal, pre, post-test design Own matched pairs with measures undertaken pre, during and post SA stay. Participation voluntary.	ERASMUS L1 English learners of French or Spanish spending a compulsory academic year living in a French or Spanish speaking country	Oral fluency measured via two speaking tasks which were both picture-based narratives Oral proficiency measured via an elicited imitation test where individuals had to repeat stimuli orally as accurately as possible	25 L2 learners of French; mean age at start of study was 20; previous experience averaged 11 years 24 L2 learners of Spanish; mean age at start of study was 21; previous experience averaged 6 years	General proficiency, mean syllable duration, mean silent pause, no. of silent pauses per second all saw sig. increase in both groups Corrections per second: Sig. increase Spanish; decrease French Filled pauses: Sig increase French; decrease Spanish. Length of silent pause and no. of repetitions per seconds no sig. difference over time in each group.

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017b). Anglophone Students Abroad ²	Longitudinal, pre, post-test design Own matched pairs across six timepoints pre, during and post SA stay. Participation voluntary.	ERASMUS L1 English learners of French spending a compulsory academic year living in a French speaking country	Global oral proficiency as measured by a French Elicited Imitation Test Oral fluency and accuracy measured using a monological narrative task Written fluency and accuracy measured using a timed argumentative essay	29 L2 learners of French; mean age was 20; previous experience averaged 11 years; mean age of first exposure was 9.5 years old	Sig. change in general proficiency over time Oral: Sig. increase in oral fluency and accuracy Written: sig. increase in fluency but not accuracy Sig. change in lexical complexity, marginal change in syntactic complexity
Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017c). Anglophone Students Abroad ²	Longitudinal, pre, post-test design Own matched pairs across six timepoints pre, during and post SA stay. Participation voluntary.	ERASMUS L1 English learners of Spanish spending a compulsory academic year living in a Spanish speaking country	Global oral proficiency as measured by a Spanish Elicited Imitation Test Oral fluency and accuracy measured using a monological narrative task Written fluency and accuracy measured via a timed essay	27 L2 learners of Spanish; mean age was 20.5; previous experience averaged 5.5 years; mean age of first exposure was 15 years old	Sig. change in general proficiency over time Oral: Sig. increase in oral fluency and accuracy Written: Sig. increase in error-free T-units (accuracy) Sig. change in lexical complexity, marginal change in syntactic complexity
Mitchell, R., & McManus, K. (2015). Subjunctive Use and Development in L2 ²	Longitudinal, pre, post-test design Own matched pairs with measures undertaken pre, during and post SA stay. Participation voluntary.	ERASMUS L1 English learners of French spending a compulsory academic year living in a French speaking country	The use of subjunctive (grammar) measured using three measurements: an argumentative writing task, semi-structured oral interviews and a grammatical judgement task	29 L2 learners of French; mean age at start of study was 20; previous experience averaged 10.4 years	Subjunctive used more at pre-test, decreased while abroad and then increased when returning to academic study Subjunctive use more frequent in writing than speech. Little change over time in scores of the GJT.
Tracy-Ventura, N. (2017). Combining corpora and experimental data to investigate language learning during residence abroad: A study of lexical sophistication ²	Longitudinal, pre, post-test design Own matched pairs across six timepoints pre, during and post SA stay. Participation voluntary.	ERASMUS L1 English learners of Spanish spending a compulsory academic year living in a Spanish speaking country	Vocabulary size as measured via the X-lex test	27 L2 learners of French; mean age at start of study was 20.5; previous experience averaged 5.5 years	Participants sig.ly increased their knowledge of less frequent vocabulary over time Participants sig. increased their productive use of less frequent vocabulary over time

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 ^{3*}	Pre-test/post-test, quasi-experimental design. Non-random group assignment. Participation voluntary.	ERASMUS L1 German learners of English studying at an L1 English speaking university. Intervention group consists of those who stay abroad for 3 months and 9 months. Control group consists of domestic-based learners who failed in their application onto ERASMUS.	General L2 proficiency measured via a c-test	143 L2 learners of English split across three groups. Short stay ($n = 45$); Long-stay ($n = 54$); Control ($n = 44$) Mean previous experience of English: 8.69 Average starting proficiency of all groups was B2 (upper intermediate)	Both YA groups experienced sig. improvement in overall L2 proficiency across 3 months. AH group made no sig. change during this time Long-stay group maintained proficiency gains – sig. difference between T2 and T3. Ah group also made sig. gains between T2 and T3 and between group differences at T3 were not sig.
Hessel, G. (2017). A new take on individual differences in L2 proficiency gain during study abroad ³	Pre/post-test design with no AH control group. Group assignment non-random. Participation voluntary.	ERASMUS L1 German learners of English studying at an L1 English speaking university for 3 months.	General L2 proficiency measured via a c-test	96 L2 learners of English split across two groups. SA long stay ($n = 52$); SA short stay ($n = 44$). Mean previous experience of English: 8-9 years	The results of both groups were merged, and paired t-tests showed a sig. change across the 3 months suggesting SA is a beneficial context for language learning
Hessel, G., & Vanderplank, R. (2018). What difference does it make? Examining English proficiency gain as an outcome of participation in ERASMUS study abroad programmes in the UK ^{3*}	Pre-test/post-test, quasi-experimental design. Non-random group assignment. Participation voluntary.	ERASMUS L1 German learners of English studying at an L1 English speaking university. Intervention group consists of those who stay abroad for 3 months and 9 months. Control group consists of domestic-based learners who failed in their application onto ERASMUS.	General proficiency as measured via a C-Test	136 L2 learners of English split across three groups. Short stay ($n = 44$); Long-stay ($n = 52$); Control ($n = 40$) Average previous learning experience = 8.69 of domestic formal instruction. All at B2 (Upper intermediate) at the beginning of the intervention.	Both YA groups experienced sig. improvement in overall L2 proficiency across 3 months. AH group made no sig. change during this time Long-stay group-maintained proficiency gains – sig. difference between T2 and T3. AH group also made sig. gains between T2 and T3 and between group differences at T3 were not sig.
Howard, M. (2005). Second language acquisition in a study abroad context: A comparative investigation of the effects of study abroad and foreign language instruction on the L2 learner's grammatical development ⁴	Cross sectional, comparison (SA vs AH vs AH) group design. Non-random group assignment. Participation voluntary.	ERASMUS L1 English learners of French either living SA or AH Intervention group consists of those who stay abroad for and 9 months. Control group consists of two types of domestic learner: a) pre-sojourners in their 2 nd year of study b) learners who forfeited the SA period to stay AH	Sociolinguistic competence (grammar): use of verb morphology for the expression of past time as measured during an interview	18 L2 learners of French split across three groups. SA ($n = 6$); AH - pre-sojourn ($n = 6$); AH – forwent SA ($n = 6$) Previous learning experience ranged between 7 and 9 years Learners were in their early 20s	SA shown to be more beneficial than AH instruction shown by the SA group using more past-time markers

Howard, M. (2006). The expression of number and person through verb morphology in advanced French interlanguage ⁴	Cross sectional, comparison (SA vs AH) group design. Non-random group assignment. Participation voluntary.	ERASMUS L1 English learners of French either living SA or AH Intervention group consists of those who stay abroad for and 9 months. Control group consists of learners who forfeited the SA period to stay AH	Sociolinguistic competence (grammar): use third person plural form, measured by means of an interview.	12 L2 learners of French split across two groups. SA ($n = 6$); AH – forwent SA ($n = 6$) Previous learning experience ranged between 7 and 9 years Learners were in their early 20s	SA group produce the third person plural more accurately than AH group suggesting increased development Much individual variation - third person plural is an area where increased development still remains to be made
Howard, M. (2008). Morpho-syntactic development in the expression of modality: The subjunctive in French L2 acquisition ⁴	Cross sectional, comparison (SA vs AH vs AH) group design. Non-random group assignment. Participation voluntary.	ERASMUS L1 English learners of French either living SA or AH Intervention group consists of those who stay abroad for and 9 months. Control group consists of two types of domestic learner: a) pre-sojourners in their 2 nd year of study b) learners who forfeited the SA period to stay AH	Sociolinguistic competence (grammar): use of subjunctive measured by means of an interview.	18 L2 learners of French split across three groups. SA ($n = 6$); AH - pre-sojourn ($n = 6$); AH – forwent SA ($n = 6$) Learners were between 20 and 22	Subjunctive form applied more in learners who had naturalistic exposure through study abroad (group 3) and through formal instruction (group 2) Much ind. variations in the group Subjunctive difficult to learn and despite living abroad, for some, the subjunctive still alludes them – SA doesn't aid the acquisition of the subjunctive form
Howard, M., Lemée, I., & Regan, V. (2006). The L2 acquisition of a phonological variable: the case of /l/ deletion in French ⁴	Cross sectional, comparison (SA vs AH) group design. Non-random group assignment. Participation voluntary.	ERASMUS L1 English learners of French either living SA or AH Intervention group consists of those who stay abroad for and 9 months. Control group consists of learners who forfeited the SA period to stay AH	Sociolinguistic competence (grammar): use of third person plural form as measured during an interview.	19 L2 learners of French split across three groups. SA ($n = 15$); AH – forwent SA ($n = 4$) Previous learning experience ranged between 5 and 6 years Learners ranged between 19 and 21	Structured L2 speakers almost never delete /l/ prior to stay in native country The speakers delete considerably more after SA period but much less than native speakers Much variation in L2 speakers
Klapper, R., & Rees, J. (2003). Reviewing the case for explicit grammar instruction in the university foreign language learning context ⁵	Longitudinal, pre, post-test design Own matched pairs across six timepoints pre, during and post SA stay. Group assignment non-random. Participation voluntary.	ERASMUS L1 English learners of German spending a compulsory academic year living in a German speaking country	Grammatical accuracy and general proficiency measured via a Grammar Test and a C-test	57 L2 learners of German. Comparable at baseline, aged between 18 and 19 at post-test. 75% of learners were language specialists with classes focused on grammatical form over meaning (FonFs); 25% of learners were not language specialists and classes focused on meaning rather than grammatical form (FonF)	AH: FonFs made statistically greater gains on C-test and grammatical test than FonF group SA: FonF group made greater progress whilst abroad in both measures than FonFs group

Rees, J., & Klapper, R. (2007). Analysing and evaluating the linguistic benefit of residence abroad for UK foreign language students ⁵	Longitudinal, pre, post-test design Own matched pairs across six timepoints pre, during and post SA stay. Group assignment non-random. Participation voluntary.	ERASMUS L1 English learners of German spending a compulsory academic year living in a German speaking country	Grammatical accuracy and general proficiency measured via a Grammar Test and a C-test	57 L2 learners of German All had comparable learning experiences, baseline proficiency and were aged between 18 and 19 at the start of the study. Learners split between programme type: 75% of learners were language specialists with classes focused on grammatical form over meaning; 25% of learners were not language specialists and classes focused on meaning rather than grammatical form	The sample made a mean progress gain of 9.47% on the C-test over the SA period. These gains were sig.ly larger than the gains made during the first two years at-home The sample made a gain of 11.19 points on the grammar test during the SA period and the difference in mean gain between AH and SA was sig.
Llanes, A., & Serrano, R. (2011). Length of stay and study abroad: language gains in two and versus three months abroad.	Pre/post-test design with no AH comparison group Voluntary participation in study assumed	ERASMUS L1 Spanish learners of English spending 2/3 months living in an English-speaking country	Written fluency, accuracy and complexity measured via a writing task Oral fluency, accuracy and complexity via narrative picture task	46 L1 Spanish/Catalan learners of English Age range between 19 and 33 21 learners stayed abroad for 3 months and 25 stayed abroad for two months (in the context of this study)	No sig. differences in gain scores between the two groups in any of the oral or written measures tested.
Serrano, Llanes, A., & Tragant, E. (2011). Analysing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs study abroad in Europe ^{6*}	Pre-test/post-test, quasi-experimental design. Group assignment non-random. Participation voluntary.	ERASMUS L1 Spanish learners of English spending time either abroad or at-home Intervention group consists of those who stay abroad for 3 months. Control group consists of two types of domestic learner: a) semi-intensive AH b) intensive AH – differ on contact hours.	Oral and written complexity, accuracy and fluency Writing: Produce two descriptive essays (one pre; test, one post-test) with learners given 15 minutes to complete task Oral: Picture-elicited narrative task	131 L2 learners of English split across three groups. SA (<i>n</i> = 25); AH intensive (<i>n</i> = 69); AH semi-intensive (<i>n</i> = 37)	No differences between intensive AH and SA groups on writing and oral measures Compared to AH semi-intensive, SA group sig.ly developed more in written and oral productions in terms of fluency and lexical complexity

Llanes, A., Tragant, E., & Serrano, R. (2012). The role of individual differences in a study abroad experience: the case of ERASMUS students ⁶	Pre/post-test design with no AH control group. Group assignment non-random. Participation voluntary.	ERASMUS L1 Spanish learners of English spending 3 months living in an English-speaking country	Oral and written complexity, accuracy and fluency Writing: Produce two descriptive essays (one pre-test, one post-test) with learners given 15 minutes to complete task Oral: Picture-elicited narrative task	24 L2 learners of English; age range between 19 and 24; previous experience averaged 13 years	Sig. increase in written fluency, oral fluency and oral lexical complexity
Serrano, R., Tragant, E., Llanes, A. (2012). A Longitudinal Analysis of the Effects of One Year Abroad ⁶	Pre/post-test design with no AH control group. Group assignment non-random. Participation voluntary.	ERASMUS L1 Spanish learners of English spending an academic year living in an English-speaking country	Oral and written complexity, accuracy and fluency Writing: Produce three descriptive essays with learners given 15 minutes to complete task Oral: Picture-elicited narrative task	14 L2 learners of English; age range between 20 and 24; began learning English at ages 6-10	Oral fluency, lexical richness and accuracy sig.ly improve over the year abroad One semester long enough to see sig. improvement in oral fluency and lexical richness. Written fluency, syntactic complexity, lexical richness and accuracy sig.ly improved across the year One semester not enough to see a sig. change in any of the written measure
Llanes, A., & Munoz, C. (2013). Age Effects in a Study Abroad Context: Children and Adults Studying Abroad and at-home ⁶ *	Pre-test/post-test, quasi-experimental design. Group assignment non-random. Participation voluntary.	ERASMUS L1 Spanish learners of English spending 2 or 3 months living in an English-speaking country	Oral and written complexity, accuracy and fluency Writing: Produce two descriptive essays (one pre-test, one post-test) with learners given 15 minutes to complete task Oral: Picture-elicited narrative task	66 L2 learners of English split across two contexts SA ($n = 46$); AH intensive ($n = 20$) Mean age of all adults = 20.9; average age of onset = 8.42 and all had received over 1,620 hours of formal instruction	SA context more beneficial than the AH context for the improvement in oral skills but no real difference in writing skills
Barron (2019). Using corpus-linguistic methods to track longitudinal development: Routine apologies in the Study Abroad context	Pre/post-test design with no AH control group. Group assignment non-random. Participation voluntary.	ERASMUS L1 English learners of German studying at an L1 German speaking university. Sojourners remain abroad for 10 months	Using apologies (pragmatics) measured via a discourse completion task	33 L2 learners of German Average age at start of SA was 19.3 Previous experience in native German speaking environment was between zero and six months.	Three directions of change recorded Some apologies remained stable across the year (e.g., use of explicit apologies Some apology behaviours were recorded towards the L2 native norm. Some apology behaviours were recorded as having a non-linear development away from the L2 native norm

Ife, A., Vives Boix, I., Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups	Pre/post-test design with no AH control group. Group assignment non-random. Participation voluntary.	ERASMUS L1 English learners of Spanish studying at an L1 Spanish speaking university. Intervention group abroad for 3 months and 6 months.	Vocabulary size measured via two measures: A translation task and word selection task	36 L2 learners of Spanish with all learning Spanish at degree level 21 were B2 and 15 were C1 (advanced) at the beginning of the intervention. Learning age and previous experience no	Sig. change across time with individuals performing better on the translation task than the misfit task. At T2, individuals sig.ly knew more words and performed better on the tasks No evidence to suggest that intermediate learners improve more so than advanced learners across the year
Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language	Pre/post-test design with no AH control group. Group assignment non-random. Participation voluntary.	ERASMUS Various European L1's using English at an L1 English speaking university for a period of six months	Vocabulary size measured using the Eurocentres Vocabulary Size Test	53 L2 users of English most of whom were not majoring in English (majoring in management science)	Pre-post difference in vocabulary size is sig. ($p = <.001$) When comparing mean change at-home and abroad, when abroad, the average growth rate is about four times as big and sig. ($p = <.001$) There were however substantial individual differences in growth rates.
Regan (1995). The Acquisition of Sociolinguistic Native Speech Norms	Pre/post-test design with no AH control group. Group assignment non-random. Participation voluntary.	ERASMUS Learner were abroad for an academic year. L1 English learners of French studying at an L1 French speaking university.	Sociolinguistic competencies (grammar) measured during an interview and then each instance of negation 'ne' was coded	6 L2 learners of French	In relation to negation, study abroad makes almost no difference. Students post SA, use the negation more but they have not yet learnt the precise deletion rules as applied by native speakers
Li, L. (2014). Language Proficiency, Reading Development, and Learning Context *	Pre/post-test, quasi-experimental design. Group assignment non-random. Participation voluntary.	Non-ERASMUS L1 English learners of Chinese Intervention group stays abroad for eight weeks. Control group remains in domestic formal instruction (1 semester)	Chinese language global proficiency test capturing components of listening, grammar, reading, translation and writing Reading comprehension test: 10 multiple choice questions	73 L2 learners of Chinese split across three proficiency groups. SA group ($n = 35$); mean age 20.7. AH group ($n = 38$); mean age 21.4	Proficiency: Beginners showed no change in either contexts. Sojourners in Intermediate and Advanced groups sig.ly improved more than AH group Reading: Sig. difference in Intermediate group only

Jochum, C. (2014). Measuring the Effects of a Semester Abroad on Students' Oral Proficiency Gains: A Comparison of At-home and Study Abroad *	Pre/post-test, quasi-experimental design. Group assignment non-random. Purposive sampling based on a set criterion with students volunteering to partake.	Non-ERASMUS L1 English learners of Spanish studying at a L1 Spanish speaking university. Intervention group stays abroad for 3 months. Control group remains in domestic formal instruction	Oral proficiency as measured by the Oral Proficiency Interview (OPI)	18 L2 learners of Spanish. SA group ($n = 9$); AH group ($n = 9$) All had completed 1 – 8 semesters of Spanish study with mean proficiency at Intermediate-low for both groups.	SA group improved their oral proficiency more than those AH and at post-test sig. between-group differences were found.
Segalowitz & Freed (2004). Context, contact, and cognition in oral fluency acquisition*	Pre/post-test, quasi-experimental design. Group assignment non-random. Purposive sampling based on a set criterion with students volunteering to partake.	Non-ERASMUS L1 English learners of Spanish studying at a L1 Spanish speaking university. Intervention group stays abroad for 3 months. Control group remains in domestic formal instruction	Oral proficiency: Oral Proficiency Interview (OPI) Cognitive tasks include attention control and lexical access	40 L2 learners of Spanish. SA group ($n = 22$); mean age 20.68 AH group ($n = 18$); mean age 23.39	SA students made sig. gains in five of the eight oral measures – OPI, Turn, Rate, Filler-free, and Fluent-run; students in the AH did not show sig. gain in any of these measures. Interaction effect for Turn; Rate; and Filler-free suggesting that the SA changed sig.ly more than AH group

* Included in the in-depth review

^{1,2,3} Studies use the same sample

Appendix E: PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	57
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	N/A
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	57-60
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	63
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	63
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	66
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	74
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	70
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	74-75
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	74-75
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	N/A
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	75
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I ²) for each meta-analysis.	N/A
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	75
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	87-89
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	89-90
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	99-106
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	104-105
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	106-107
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	106-107
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	10

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097.



1. Email

What is your email address? (I recommend using your university email address)

2. Consent

You have been invited into a research study which investigates fluctuations of thoughts, feelings and behaviours among sojourners and non-sojourners, together with linguistic change in those who study a modern foreign language. The study is conducted by Mr Ian Moore from the School of Education, Durham University, Durham.

This is the first questionnaire you will be asked to complete. By clicking the "Yes, I agree" button you agree to participate in the study and confirm the following:

- I agree to participate in this study, the purpose of which is to capture changes in one's personality and linguistic development.
- I have read the participant information sheet and understand the information provided.
- I have been informed that I may decline to answer any questions or withdraw from the study without penalty of any kind.
- I have been informed that all of my responses will be kept confidential and secure, and that I will not be identified in any report or other publication resulting from this research.
- I have been informed that the investigator will answer any questions regarding the study and its procedures. Mr Ian Moore, School of Education, Durham University can be contacted via email: ian.j.moore@durham.ac.uk

By clicking "No, I do not agree" you will be screened out of the study and no information will be collected.

Any concerns about this study should be addressed to the School of Education Ethics Sub-Committee, Durham University via email to ed.ethics@durham.ac.uk

Thank you for your participation.

- Yes, I agree
 No, I do not agree

Introduction

Thank you for taking part in this study. Your participation really is appreciated. The following questionnaire will be conducted only once. You will be asked to create a personal ID. You will be asked this each time you complete a survey. This will ensure you remain anonymous across all data-collection points and enables data collected to be matched.

This questionnaire will take roughly 5 minutes to complete.

3. Personal ID

Your personal ID must be made up of the following information:

Month born (Please spell in letters)

Number of older siblings (if none, write X)

First two letters of your birth place

First letter of mother's/guardian's name

For example: JULYXTAC

Please create a personal ID.

4. What is your gender?

- Male
- Female
- Prefer not to say

5. How old are you?

6. What is your mother tongue?

7. Which university do you currently attend?

8. Which languages do you learn at degree level?

- French
- German
- Italian
- Spanish
- None
- Other

9. Please state the full title of your honours degree.

10. In the academic year 2018/19, will you be a 2nd year or 3rd year student?

- 2nd year
- 3rd year

11. Are you abroad or staying at home this academic year (2018-2019)?

- Living abroad
- Staying home

12. Host country?

In which country (or countries) will you be completing your year abroad?

13. In which context(s) will you complete your year abroad?

- Studying
 Teaching
 Working
 Studying and working (split)
 Teaching and studying (split)
 Teaching and working (split)

Other

14. How long will you be staying in your chosen country?

- 1 semester (dividing the year abroad into two countries)
 The whole academic year (staying in the same country the whole year)

Other

15. Where will you be staying? (accommodation type)

- Student accommodation (with native speakers of your year abroad country e.g. German)
 Student accommodation (with other Erasmus students)
 Student accommodation (with native speakers of your first language e.g. English)
 Private shared accommodation (with native speakers of your year abroad country e.g. German)
 Private shared accommodation (with other ERASMUS students)
 Private shared accommodation (with native speakers of your first language e.g. English)
 Living alone
 A homestay (living with a host family)
 I'm currently not sure

Other

16. Are you required to complete assessed work whilst abroad? If yes, what are you required to do?

Yes

No

17. What is your expected start date of the year abroad?

18. What is the expected end-date of your year abroad?

If you do not know the date exactly just put the month or state 'I do not know'

19. Current GPA? (If known)? This is a percentage score.

Please leave blank if you not wish to say.

What is your current GPA for speaking?

What is your overall current GPA?

20. Do you plan on joining language classes whilst abroad?

21. How many months are you spending abroad in total?

22. Which academic field are you in?

- Languages and Linguistics
- Arts and Humanities
- Sciences
- Mathematics
- Accounting and Economics
- Education
- Business
- Politics
- Journalism
- Psychology
- Sociology
- Art and Design

Other

23. What was your highest A-level?

- A*
- A
- B
- C
- D
- D – U
- Prefer not to say

24. What was your lowest A-level?

- A*
- A
- B
- C
- D
- D – U
- Prefer not to say

25. How any siblings do you have?

- None
- 1
- 2
- 3
- 4
- 5 or more
- Prefer not to say

26. Have you ever lived abroad for more than one month in a row prior to this academic year (2018/19)?

- Yes
- No

Previous experience abroad

Please add up how many months you have spent abroad prior to the year abroad

Your answers have been submitted. I recommend you add my email address to your VIP list. This will ensure you do not miss any emails from me. Please copy the address <https://www.techlicious.com/how-to/never-miss-an-important-email-on-your-smartphone/> to see instructions on how to set up a VIP email if you wish to do this. I have now way of knowing if you have done this. You may now close the browser window or tab.

If you would like to contact me, please click [here](#)



Trait Personality

This questionnaire is aimed at measuring your general thoughts, feelings and behaviours. There is no right or wrong answer and please answer truthfully.

This questionnaire will take roughly 5 minutes to complete.

Personal ID

What is your personal ID?

Month born; Number of older siblings (if none, write X); First two letters of your birth place; First letter of mother's/guardian's name. For example: JULYXTAC

Here are a number of characteristics that may or may not apply to you. Please rate yourself according to the adjectives (disagree strongly – agree strongly) seen on the top line to indicate the extent to which you agree or disagree with that statement.

I see Myself as Someone Who ...

Disagree strongly Disagree a little Neither agree nor disagree Agree a little Agree strongly

Is talkative

Tends to find fault with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does a thorough job	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is depressed, blue	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is original, comes up with new ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is reserved	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is helpful and unselfish with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can be somewhat careless	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is relaxed, handles stress well	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is curious about many different things	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is full of energy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Starts quarrels with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is a reliable worker	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can be tense	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is ingenious, a deep thinker	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Generates a lot of enthusiasm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Has a forgiving nature	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tends to be disorganized	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Worries a lot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Has an active imagination	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tends to be quiet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is generally trusting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tends to be lazy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is emotionally stable, not easily upset	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is inventive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Has an assertive personality	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can be cold and aloof	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Perseveres until the task is finished	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Can be moody	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Values artistic, aesthetic experiences	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Is sometimes shy, inhibited	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is considerate and kind to almost everyone	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does things efficiently	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Remains calm in tense situations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Prefers work that is routine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is outgoing, sociable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is sometimes rude to others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Makes plans and follows through with them	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gets nervous easily	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Likes to reflect, play with ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Has few artistic interests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Likes to cooperate with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is easily distracted	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Is sophisticated in art, music, or literature	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1. Do you think you have changed as a person over the past 3 months

- Yes
- No
- Not sure
- Not applicable

Thank you for your participation

If you have any comments or suggestions regarding this test, please do not hesitate contacting the researcher at ian.j.moore@durham.ac.uk.

You will be asked to complete another trait personality test in the middle and at the end of the academic year.

Thank you for supporting this study.

Please press next to submit your results.



Personality facets

The aim of this questionnaire is to capture how your general thoughts feelings and behaviours over the past month. The questionnaire will also ask for some detail regarding the situations you have been in.

This questionnaire will take roughly 10 minutes to complete and you will be asked to complete it only once a month.

Please complete this questionnaire within 4 days of receiving it.

1. Personality facets

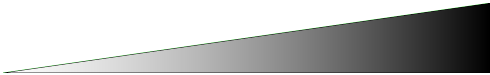
Personal ID:

Month born; Number of older siblings (if none, write X); First two letters of your birth place;
First letter of mother's/guardian's name. For example: JULYXTAC

2. Rate the statements below for how accurately they reflect the way you have been generally feeling and behaving over the past month. Do not rate what you think you should do, or wish you do, or things you no longer do. Please be as honest as possible.

Well being

In the last month ...

	Not at all	Only occasionally	Sometimes	Often	Most/All of the time
					
I have felt tense anxious or nervous	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt I have someone to turn to when things go wrong	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt OK about myself	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt able to cope when things go wrong	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have been troubled by aches, pains or other physical symptoms	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have been happy with the things I have done	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have had difficulty getting to sleep or staying asleep	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt warmth or affection for someone	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have been able to do most things I needed to	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt criticised by other people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt unhappy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have been irritable when with other people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt optimistic about my future	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have achieved the things I wanted to	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. Interaction

Please think of a typical week in the past month and indicate the amount of time (in hours) you have spent using your second language (L2) in the following activities. Your L1 is your native language.

Communicating with native L2 speakers (e.g. German)

Listening to L2 radio and music

Writing in the L2 e.g. emails

Watching L2 TV/films/streaming videos

Reading L2 material e.g. books/newspapers

Communicating with native L1 speakers (e.g. English)

Listening to L1 radio and music

Writing in the L1 e.g. emails

Watching L1 TV/films/streaming videos

Reading L1 material e.g. books/newspapers

4. Resiliency

In the last month ...

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I have tended to bounce back quickly from hard times	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have had a hard time making it through stressful events	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It hasn't taken me long to recover from a stressful event	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have found it hard to snap back when something bad happens	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have usually come through difficult times with little trouble	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have tended to take a long time to get over the set-backs experienced	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. Anxiety

In the last month ...

	Almost never	Infrequently	Sometimes	Often	Almost always
I have felt calm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt tense	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt upset	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt nervous, worried or anxious	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt relaxed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have felt content	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Curiosity

	Very Slightly or Not At All	A Little	Moderately	Quite a Bit	Extremely
I have actively sought as much information as I can in new situations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have enjoyed the uncertainty of everyday life	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have been at my best when doing something that is complex or challenging	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have searched for new things or experiences	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have enjoyed doing things that are a little frightening	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have searched for experiences that challenge how I think about myself and the world	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have preferred jobs that are excitingly unpredictable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have frequently sought out opportunities to challenge myself and grow as a person	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have embraced unfamiliar people, events and places	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have viewed challenging situations as an opportunity to grow and learn	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. Where have you spent the most time in August – home or abroad? Take the number of days you have spent in each and select the context which has the highest number.

8. Have you created a close bond with anybody (friendship or otherwise) in the past 2 weeks. If yes, how many?

Yes

No

9. Take a weekly average: how frequently have you been in contact with friends and family from home in the past 4 weeks?

- Every day
- Every second day
- Twice a week
- Once a week
- Not at all
- I currently live at home (outside term time)

10. Has a particular positive event occurred?

- Yes
- No
- Prefer not to say

11. Type of positive event? Please leave blank if no specific positive event occurred.

- Went to a concert/festival
 - Explored a new area/country with friends/alone
 - Visited Family/friends/partner or came to visit
 - Overcame a new challenge
 - Making new friends
 - Started a new relationship
 - Sporting success
 - Academic success
 - Went on a date
 - Celebration party
 - Went home to visit
 - Personal success
 - Completed a new challenge
 - Physical/mental health improvements
 - Job offer/success
 - Prefer not to say
 - Other
-

12. Has a particular negative event occurred in your life in the past 4 weeks? If yes, please complete the next question too.

- Yes
- No
- Prefer not to say

13. If yes, please select from the options below; if no, please leave blank:

- Death or illness of a family member or friend
- Personal injury
- Accommodation difficulties
- Lost passport/important document/lost item
- Broke up from a relationship
- Friendship/Relationship difficulties
- Financial Issues
- Job rejection
- Isolation/Lonliness
- Attacked
- Academic application rejection e.g. Masters
- Internship issues
- Administration issues
- Poor mental health
- Family arguments
- Physical health issues
- Academic issues
- Poor exam result
- Prefer not to say
- Other
-

14. Take a weekly average: how frequently have you socialised with native speakers outside your specific context e.g. studying, teaching or working?

- Frequently (every day)
- Often (every 2 days)
- Sometimes (every 3 days)
- Rarely (every 4 days or more)

15. Take a weekly average: how frequently have you found yourself becoming bored in the past 4 weeks?

- Frequently (every day)
- Often (every 2nd/3rd day)
- Sometimes (every 4th/5th/6th day)
- Rarely (once a week)
- Not at all

16. Are you an active member in any clubs or societies? If yes, what type(s)? (e.g. Sport – Football) Please give up to four examples.

- Yes
- No

17. Do you feel you play an active social role in the local community?

- Yes
- No

18. How many times have you been home (family home) in the past 4 weeks?

- Zero
- Once
- Twice
- Thrice
- More than three times
- I currently live at home (Outside term time)

19. Have you visited any other countries in the past 4 weeks? (apart from your home or host country)? If yes, how many?

- Yes
- No

20. Do you feel welcomed in the local community? (Please complete if home or abroad)

- Yes
- No

21. Have you felt predominantly lonely, culture shocked, or homesick in the past 4 weeks? (please answer whether home or abroad)

Yes

No

Thank you for your participation

If you have any comments or suggestions regarding this test, please do not hesitate contacting the researcher at ian.j.moore@durham.ac.uk.

You will be asked to complete this test again once a month.

Please press next to submit your results.

Last Page

Your answers have been submitted and you may now close the browser window or tab.

If you would like to contact me, please click [here](#)



The following questionnaire is a short snapshot of your **CURRENT** thoughts, feelings and behaviours. Please rate how the following adjectives reflect these thoughts. You will also be asked about your current situation. This questionnaire will take 2 minutes to complete.

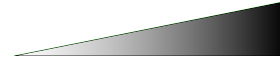
What is your Personal ID?

Month born (in letters); Number of older siblings (if none, write X); First two letters of your birth place; First letter of mother's/guardian's name. For example: JULYXTAC

1. What is the current time? (in your specific country)

How do these adjectives reflect your current thoughts, feelings and behaviours?

Not at all Completely



Uncreative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rude	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Talkative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Insecure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Assertive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Curious	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quiet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lazy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Forgiving	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Artistic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Calm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hard-working	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intellectual	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Anxious	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trustful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Energetic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cooperative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Emotional	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unreliable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Relaxed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Responsible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cold	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Organised	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unimaginative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unadventurous	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

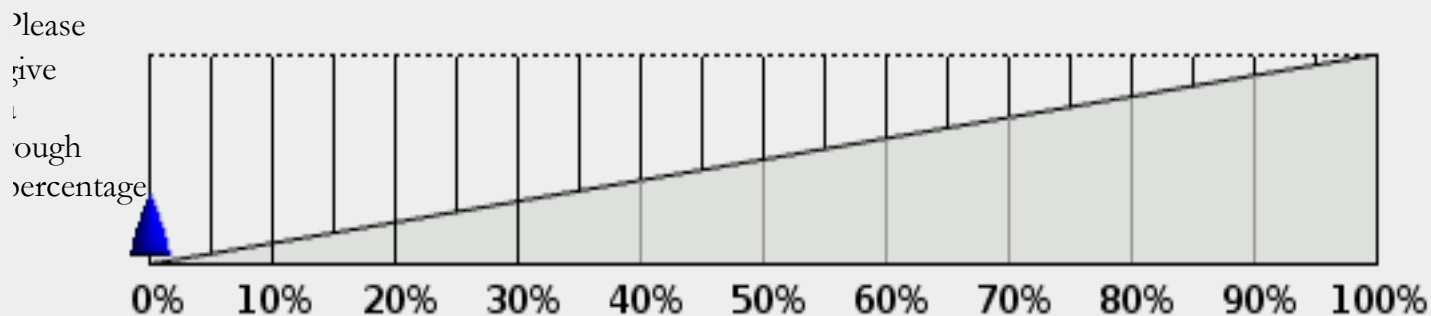
ersons

How many people are you currently with?

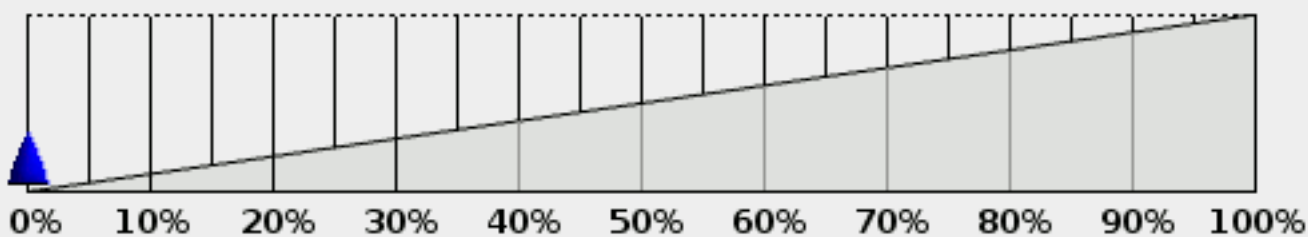
[Please choose]



From your response to the previous question. How many of these are native speakers of your learnt second language (L2)? For example, if French is your L2 – how many people around you are French native speakers? Please leave BLANK (0) if you are NOT a language learner.



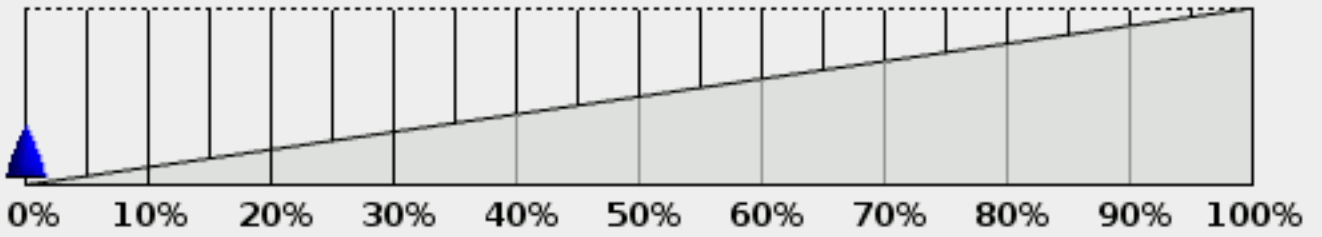
On a scale of 0-100%, how familiar are you in general with the people currently around you? (Please leave as 0 if alone)



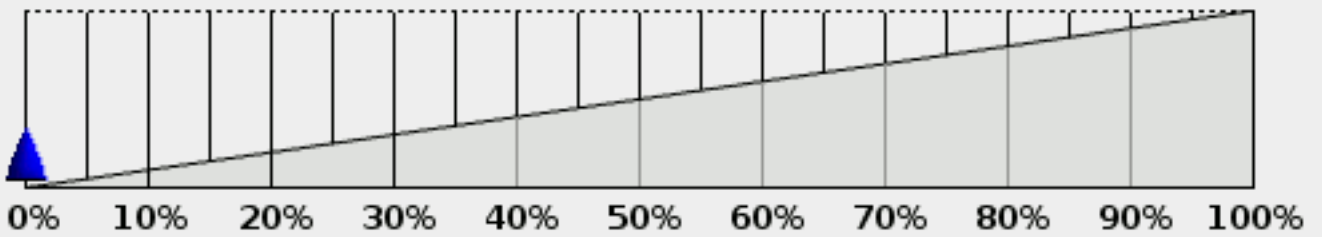
Location

Where are you currently?

On a scale of 0-100%, how familiar are you with your current environment?



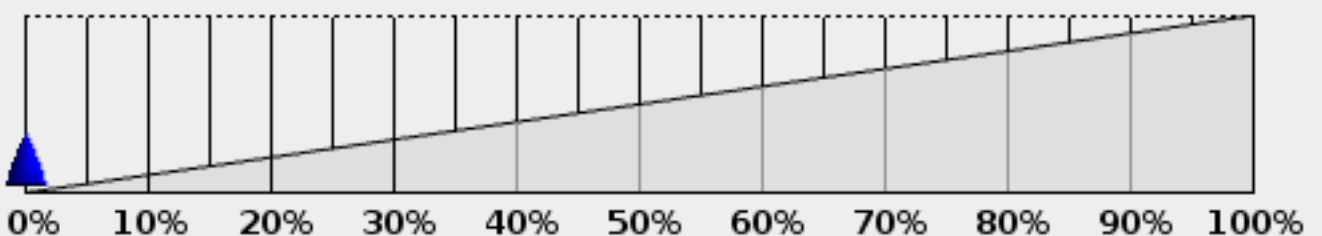
On a scale of 0-100%, how secure do you feel in your current environment?



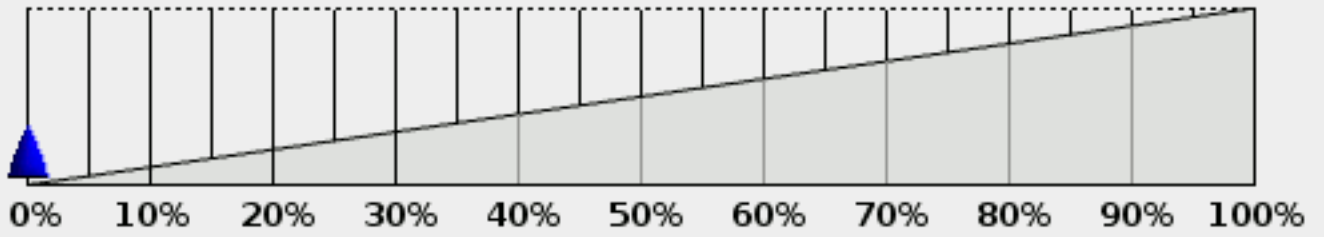
. What are you currently doing?

'ask

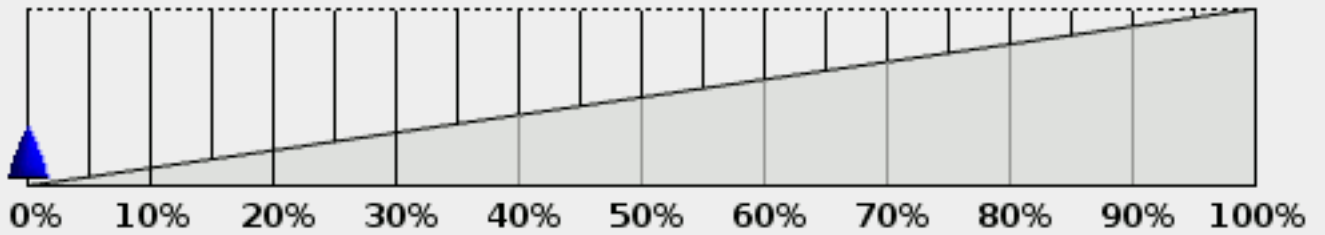
On a scale of 0-100%, is the current task you are currently undertaking enjoyable?



On a scale of 0-100%, is the current task you are currently undertaking familiar to you?



On a scale of 0-100%, does the current task you are undertaking benefit you in any way personally?



On a scale of 1 to 7, please state how much your current situation reflects these statements

	Not at all	Hardly	Not really	Maybe	Quite	Mostly	Totally
Work has to be done	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deep thinking is required	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Somebody is being threatened, accused, or criticized	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Potential romantic partners are present	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The situation is pleasant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The situation contains negative feelings (e.g., stress, anxiety, guilt, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Somebody is being deceived	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Social interactions are possible or required	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using the second language (e.g., French) is required (speaking, reading, listening, writing)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Thank you for your participation

If you have any comments or suggestions regarding this test, please do not hesitate to contact the researcher at ian.j.moore@durham.ac.uk.

Thank you for supporting this study.

Please press next to submit your results



1. German C-Test

You have been given this test because you selected this language in the demographic questionnaire. If you have been given the wrong test please contact ian.j.moore@durham.ac.uk.

On the next page is a set of instructions.

There may not always be one right solution and a C-test is designed so that even a native speaker may only score between 80-95%.

This test will take roughly **20 minutes** to complete and it is expected that each question will take **5 minutes** to complete.

Instructions

To complete this test please fill in the gaps - **you are required to fill in the second half of every second word**. For example;

He li_____ green app_____

He **likes** green **apples**.

The infor_____ was unc_____

The **information** was **unclear**.

You can either have the same number of letters in the gap as come before it or 1 more. However, **having 2 more letters or 1 less is not permitted**

e.g. Po? = Po**ke** - 2+2 = OK

= Po**ker** - 2+3 = OK

= Po**t** = 2+1 = X

Pay close attention to spelling, verb tenses and apostrophes (') - these count as a letter in the rules described above.

You do not lose points for wrong answers so it is **better to have an answer** than an empty gap but if you do not know an answer please place a **dash (-)** in the box.

Points will only be awarded for correct answers - any spelling or grammatical mistakes will mean that the points are lost.

The boxes **ALWAYS** relate to the letters which come before the box; **NEVER** after.

Personal ID

What is your personal ID?

Month born; Number of older siblings (if none, write X); First two letters of your birth place; First letter of mother's/guardian's name. For example: JULYXTAC

2. Hiddensee

Hiddensee ist eine kleine Insel in der Ostsee. Vielleicht geh diese In zu ei

der let Idyllen i Deutschland mit herrl Landschaft, kla
 Wasser, saub Luft. We es wa ist, ka
 man si

in d kühlen Ostsee abkü . Auf Hiddensee gi es g

keine Indu . Nur Pers , die beru unbedingt e

Auto benö , dürfen a der In Auto fah

Alle and fahren Fahrrad oder gehen zu Fuß. Hiddensee ist ein Ferienparadies, dass das ganze
Jahr von sehr vielen Urlaubern besucht wird.

3. Vitamine

Vitamine spenden dem Körper Gesundheit und Fitness. Besonders i Herbst u
Winter i bei vie Menschen d Befürchtung we
verbreitet, i dieser Ze zu we Vitamine aufzu

. Deshalb bie Lebensmittelherst eine Vielza von Prod

an, de Vitamine zuge wurden. E gibt z.B. versch
 Multivitaminmarmeladen od Vitaminsäfte. Vi glauben, da

Nahrungsmittel m Vitaminzusätzen se gesund si . Sie ge
 dafür sehr viel Geld aus. Aber frisches Obst oder Gemüse zu kaufen ist besser.

4. Benzinpreise

Die hohen Benzinpreise, die teureren Flugtickets, die Inflationsraten und nicht zuletzt die ständige Ermahnung, den Gürtel enger zu schnallen, da schwere Zeiten auf uns zukommen - dies ist [] nach Ansicht

[]

der Reiseveranstalter [] die Gründe [] dafür, dass [] eingefleischte Fernreisefans in [] Urlaubsstrategie geändert haben und [] im Extremfall []

auch bereit sind, in [] deutschen Ländern [] Ferienzeiten

zu machen. Man [] bundesdeutsche Ferienort [] haben die []

Trend rechtzeitig [] erkannt, investieren [] in die [] Sommermonate

[]

als bis [] in der [] Werbung und [] stellen nicht [] Ende Juli schon fest, dass das "absolut gut angelegte Geld war".

5. Dunkelheit

Im Dezember wird es schon zeitig dunkel. Dann werden [] unzählige Lampen []

eingeschaltet, die [] in vielen [] Büros und [] Geschäften arbeiten

[]

man noch []. In vielen [] Haushalten wird [] Essen gekocht []

. Es laufen [] viele Fernseher []. Gibt es [] in die []

Momente ein [] Störung in [] einem Kraftwerk [], bekommen wir

[] trotzdem Strom [], weil Stromleitungen [] Europa verbieten [].

Dieses internationale [] Stromverbundnetz ist [] ein Beispiel [] für die

[] Zusammenwachsen der Wirtschaft in Europa. Man kann auf Störungen in der Energieversorgung so europaweit besser reagieren



1. French C-Test

You have been given this test because you selected this language in the demographic questionnaire. If you have been given the wrong test please contact ian.j.moore@durham.ac.uk.

On the next page is a set of instructions.

There may not always be one right solution and a C-test is designed so that even a native speaker may only score between 80-95%.

This test will take roughly **20 minutes** to complete and it is expected that each question will take **5 minutes** to complete.

Instructions

To complete this test please fill in the gaps - **you are required to fill in the second half of every second word**. For example;

He li_____ green app_____

He **likes** green **apples**.

The infor_____ was unc_____

The **information** was **unclear**.

You can either have the same number of letters in the gap as come before it or 1 more. However, **having 2 more letters or 1 less is not permitted**

e.g. Po? = Po**ke** - 2+2 = OK

= Po**ker** - 2+3 = OK

= Po**t** = 2+1 = X

Pay close attention to spelling, verb tenses and apostrophes (') - these count as a letter in the rules described above.

You do not lose points for wrong answers so it is **better to have an answer** than an empty gap but if you do not know an answer please place a **dash (-)** in the box.

Points will only be awarded for correct answers - any spelling or grammatical mistakes will mean that the points are lost.

The boxes **ALWAYS** relate to the letters which come before the box; **NEVER** after.

What is your personal ID?

Month born; Number of older siblings (if none, write X); First two letters of your birth place; First letter of mother's/guardian's name. For example: JULYXTAC

2. Régions

Les régions n'existent que depuis 1960. Il faut dire que pendant des siècles, la France a été un pays très centralisé. Toutes les décisions politiques étaient prises à Paris. La capitale était le centre économique et culturel le plus important de France. C'est le lien vers lequel toutes les régions étaient tournées, il n'y avait rien qui attirait tout le monde s'intéressait, et cela posait des problèmes: par exemple, il y avait trop peu d'industries en province.

3. Le vélo

Bon marché, écolo, sain... le vélo a toutes les qualités pour les Français. Ils le considèrent comme le meilleur mode de transport de l'avant-garde. Cependant, seulement 14 % des utilisateurs de vélos l'utilisent au moins deux fois par mois ! Le manque de conditions physiques, les trop longues distances à parcourir, l'absence de pistes cyclables et la peur d'accidents freinent les ambitions des cyclistes. On compte actuellement 21 millions de vélos en France (367 pour 1 000 habitants). Contre 72 millions en Allemagne (900 pour 1 000 habitants).

4. Etudier en France

Cinquième puissance économique mondiale, terre d'art et de culture, destination préférée des touristes du monde, la France figure aussi parmi les premiers pays d'accueil d'étudiants étrangers. Mais ceu [] ne chois [] pas no [] pays seul [] pour l' [], la litté [] et l [] sciences huma []. Et c [] attrait n [] date p [] d'hier. Fa [] ses étu [] en Fra [], pour beau [], c'est u [] rêve. U [] certaine en [] d'excellence, au []. Parmi l [] plus diver [] et l [] plus perfo [] du mo [], l'enseignement fran [] propose en effet une offre formidable. De plus, les droits d'inscription y sont plus que compétitifs.

5. Chiens

A l'heure où la France s'interroge sur l'attitude à adopter, plusieurs pays qui connaissent le même problème ont déjà pris des mesures rigoureuses. En Grande-Bret [], la l [] sur l [] chiens dang [] de 1991 -vo [] à l'i [] d'une sé [] d'attaques s [] des enf [] - interdit d [] vendre e [] d'élever d [] pitbulls, d [] tosas japo [], des dog [] argentins e [] des filis brési []. Leurs propri [] doivent l [] faire enreg [], leur met [] une lai [] et u [] muselière e [] public e [] souscrire une assurance spéciale. Les maîtres négligents s'exposent à six mois de prison ferme, une amende de 15 000 francs et la mise à mort de la bête.

If you have any comments or suggestions regarding this test, please do not hesitate contacting the researcher at ian.j.moore@durham.ac.uk. You will be asked to complete another language test at the end of your time in that country.

Please press next to submit your results.

Thank you for completing this test!

Your answers have been submitted and you may now close the browser window or tab.

If you would like to contact me, please click [here](#)



1. Italian C-Test

You have been given this test because you selected this language in the demographic questionnaire. If you have been given the wrong test please contact ian.j.moore@durham.ac.uk.

On the next page is a set of instructions.

There may not always be one right solution and a C-test is designed so that even a native speaker may only score between 80-95%. Please do not worry if you cannot fill in every gap.

This test will take roughly **20 minutes** to complete and it is expected that each question will take **5 minutes** to complete.

Instructions

To complete this test please fill in the gaps - **you are required to fill in the second half of every second word**. For example;

He li_____ green app_____

He **likes** green **apples**.

The infor_____ was unc_____

The **information** was **unclear**.

You can either have the same number of letters in the gap as come before it or 1 more. However, **having 2 more letters or 1 less is not permitted**

e.g. Po? = **Po**ke**** - 2+2 = OK

= **Po**ker**** - 2+3 = OK

= **Po**t**** = 2+1 = X

Pay close attention to spelling, verb tenses and apostrophes (') - these count as a letter in the rules described above.

You do not lose points for wrong answers so it is **better to have an answer** than an empty gap but if you do not know an answer please place a **dash (-)** in the box.

Points will only be awarded for correct answers - any spelling or grammatical mistakes will mean that the points are lost.

The boxes ALWAYS relate to the letters which come before the box; NEVER after.

What is your personal ID?

Month born; Number of older siblings (if none, write X); First two letters of your birth place; First letter of mother's/guardian's name. For example: JULYXTAC

2. Ladri

Pellicce, gioielli, denaro. Il classico bottino da furto in appartamento: e questa volta i ladri hanno colpito Caterina Caselli.

L'altra no [] sono ent [] nella s [] elegante ca [] nel cen [] di Mil [] e hanno av [] cinque o [] per apr [] la cassa [] di fami [] e portare v [] tutto que []

che riusc []. L'elegante cant [] era fu [] con i [] marito Piero Sugar e qua [] sono rien [] verso l'1.25 ha [] trovato l'appar []

sottosopra. I [] valore de [] refurtiva è an [] imprecisato e l' [] "Casco d'oro" ha passato tutto ieri a verificare esattamente quello che mancava.

3. Ecologia

Prendere ogni tanto l'autobus o la bicicletta. Mangiare la carne una volta in meno alla settimana.

Comprare u [] nuovo elettrod [] a ba [] consumo energ []

. Non spre [] troppa ac [] per fa [] la doc [] o qua [] si lav [] i pia []. Sono ta [] le pic [] azioni quoti [] apparentemente insigni [], ma la so [] di mil [] di que []

azioni pe [] sul bila [] ecologico glo []. Fortunatamente sta [] cambiando le abit [] di vi [] anche s [] lentamente. Tuttavia soprattutto le aziende devono fare la loro parte, cercando di ridurre il più possibile le emissioni di CO2 e gli enormi consumi.

4. Lo scandalo Parmalat

Adesso anche l'Europa deve confrontarsi con un enorme scandalo finanziario. Il ca [] Parmalat vi [] giustamente defi [] come l [] più gra [] frode finan [] della sto [] europea. S [] tratta d [] una so [] fra g [] 8 e i 10 mili [] di eu [] spariti, fo [] anche d [] più. I [] fatto c [] uno scan [] di ta [] misura co [] quello d [] Parmalat pot [] avvenire i [] Europa è st [] per mo [] uno shock perché il " vecchio Continente " era considerato finora malgrado le irregolarità contabili nei casi dell'olandese Ahold o del francese Vivendi, con il suo sistema di sorveglianza finanziaria, più sicuro rispetto al sistema capitalistico americano.

5. La cucina toscana

La cucina toscana, tipica espressione della sana e appetitosa dieta mediterranea, è apprezzata in tutto il mondo. Il motivo principale della sua celebrità è la sua genuinità e semplicità. I suoi pia [] sono ric [] di ingre [] naturali. S [] tratta d [] una cucina pov [] ma pi [] di tradi [], fondata s [] pietanze inve [] con i prod [] offerti dall'or [] e dal bo [], accompagnati d [] carni alle [] in zo [], olio d [] oliva, pa [] e vi [] da ve [] intenditori. Og [] provincia tos [] ha u [] caratteristiche b [] preciso e i su [] piatti tipici. Importanti sono le produzioni locali di salami e formaggi famosi.

If you have any comments or suggestions regarding this test, please do not hesitate contacting the researcher at ian.j.moore@durham.ac.uk. You will be asked to complete another language test at the end of the academic year.

Please press next to submit your results.

Thank you for completing this test!

Your answers have been submitted and you may now close the browser window or tab.

If you would like to contact me, please click [here](#)



1. Spanish C-Test

You have been given this test because you selected this language in the demographic questionnaire. If you have been given the wrong test please contact ian.j.moore@durham.ac.uk.

On the next page is a set of instructions.

There may not always be one right solution and a C-test is designed so that even a native speaker may only score between 80-95%. Please do not worry if you cannot fill in every gap.

This test will take roughly **20 minutes** to complete and it is expected that each question will take **5 minutes** to complete.

Instructions

To complete this test please fill in the gaps - **you are required to fill in the second half of every second word**. For example;

He li_____ green app_____

He **likes** green **apples**.

The infor_____ was unc_____

The **information** was **unclear**.

You can either have the same number of letters in the gap as come before it or 1 more. However, **having 2 more letters or 1 less is not permitted**

e.g. Po? = **Po**ke**** - 2+2 = OK

= **Po**ker**** - 2+3 = OK

= **Po**t**** = 2+1 = X

Pay close attention to spelling, verb tenses and apostrophes (') - these count as a letter in the rules described above.

You do not lose points for wrong answers so it is **better to have an answer** than an empty gap but if you do not know an answer please place a **dash (-)** in the box.

Points will only be awarded for correct answers - any spelling or grammatical mistakes will mean that the points are lost.

The boxes ALWAYS relate to the letters which come before the box; NEVER after.

2. Personal ID

What is your personal ID?

Month born; Number of older siblings (if none, write X); First two letters of your birth place; First letter of mother's/guardian's name. For example: JULYXTAC

3. Televisión

Es un fenómeno insólito. Se ha [] de é [] en e [] mercado, e [] las escu [] , en l [] oficinas, ha [] en e [] Parlamento. C [] las últ [] noticias so [] la cri [] del Atlé [] de Mad [] o el penú [] y presunto ca [] de corru [] política, l [] televisión conf [] el ma [] de l [] obsesiones nacio [] . Amada y odi [] con pas [] , no de [] a nadie indiferente. Algunos piensan que ha llegado demasiado lejos y otros esperan con curiosidad lo que queda por venir.

4. Actividades

La lista de actividades con futuro es bastante larga. Desde l [] reforestación d [] todo e [] país a rit [] forzados ha [] los cuid [] geriátricos y e [] general l [] negocios relaci [] con l [] tercera ed [] , pasando p [] la rehabil [] de vivi [] , el tur [] cultural, l [] climatización, e [] mantenimiento d [] maquinarias y ot [] cien posibi [] . Aunque par [] mentira, ha [] 30 años, u [] ingeniero rec [] licenciado po [] aspirar a trabajar en no menos de 10 empresas españolas de televisión y sonido instaladas en Barcelona.

5. El español en EE.UU.

La presencia hispánica en el territorio de los actuales Estados Unidos de América comenzó con el descubrimiento de La Florida en el año 1513. Al [] largo d [] ese si [], los espa []
 []
 explo [] la reg [] de l [] estados actu [] de Texas, Nuevo México, Arizona y Colorado. A par [] de ento [], se establ [] misiones e [] California. E [] 1821, México s [] indep [] de España. Añ [] después, ter [] la gue [] entre Estados Unidos y México, do [] México per [] la mi [] de s [] territorio. En el si [] XX, l [] oleadas migra [] de hispanoamericanos hacia Estados Unidos no sólo han rehispanizado estas zonas, sino que han introducido el idioma español en todo el país. En 1997, la cifra de hispanos se elevaba a 29,7 millones, un 11,1%.

6. Columna Conmemorativa (1968)

Se erige en el año 1968 con motivo del XIX centenario de la fundación de la Legio VII Gemina. Hasta e [] momento, l [] restos arqueol [] no hab [] desvelado q [] el antec [] del ori [] de l [] ciudad de León s [] encontraba e [] un campa [] anterior denom [] la Legio VI Victrix. P [] este mot [] , el alc [] Arroyo Quiñones apr [] la organi [] de ac [] conmemorativos d [] la fe [] del nacim [] de l [] ciudad. L [] columna, rema [] con u [] capitel corintio sobre un pedestal, se eleva en uno de los laterales de la Plaza de San Isidoro.

If you have any comments or suggestions regarding this test, please do not hesitate contacting the researcher at ian.j.moore@durham.ac.uk. You will be asked to complete another language test at the end of the academic year.

Please press next to submit your results.

Thank you for completing this test!

Your answers have been submitted and you may now close the browser window or tab.

If you would like to contact me, please click [here](#)



Perceived Competencies

The following questionnaire is aimed at measuring your perceived competencies in your second language (L2). There is no wrong or right answer and your data will be held in the strictest of confidentiality. If you learn two languages, please complete two SEPARATE questionnaires.

This questionnaire will take roughly 5 minutes to complete.

1. Personal ID

What is your personal ID?

Month born; Number of older siblings (if none, write X); First two letters of your birth place; First letter of mother's/guardian's name. For example: JULYXTAC

2. What is your L2?

- French
- German
- Italian
- Spanish

Other

3. I can statements:

Please rate yourself between 0 – 10 on how well you can do the following statements.

0 = cannot do at all

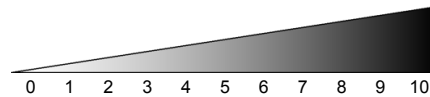
3 = have great difficulty to do well

5 = can do somewhat

7 = can do quite well

10 = can do very well

I can ...



follow conversation on familiar topics even in noisy environments (e.g. a bar)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
understand movies/TV/radio which are in a standard dialect	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
use a variety of strategies to achieve comprehension e.g. listening for main points	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
understand texts which are in my field of interest in detail	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
understand texts which offer review and/or criticism and then give a summary of the main points	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
quickly look through a manual and find a solution to a problem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
express myself clearly in everyday conversation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
construct reasoned arguments and express myself clearly and appropriately	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
initiate, maintain, and end a conversation effectively through use of fillers, turn-taking etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
write clear and detailed texts (e.g. reports) on topics related to my field	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
write a short review of a film or a book	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
develop a clear and detailed for and against arguments with clear evidence and accurate structure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Thank you for your participation

If you have any comments or suggestions regarding this test, please do not hesitate contacting the researcher at ian.j.moore@durham.ac.uk.

You will be asked to complete another competencies questionnaire at the end of your period abroad

Please press next to submit your results.

Reminder: if you study two language, please complete another survey by clicking the same link.

Appendix L: Links to Language Instruments

German

Text 1	https://www.stksachs.uni-leipzig.de/files/media/pdf/aufnahmetest/Deutschtest_02_15.pdf
Text 2	https://www.stksachs.uni-leipzig.de/files/media/pdf/aufnahmetest/Deutschtest_02_15.pdf
Text 3	https://www.uni-muenster.de/C-Test/demo/index.php
Text 4	https://www.stksachs.uni-leipzig.de/files/media/pdf/aufnahmetest/Deutschtest_02_15.pdf

French

Text 1	https://www.uni-muenster.de/C-Test/demo/index.php
Text 2	https://buchsys.zfbk.uni-giessen.de/cgi/ctest2.cgi?testcode=a1f51d94743c2eb38fd120b958ffed61
Text 3	Taken from https://buchsys.zfbk.uni-giessen.de/cgi/ctest2.cgi?testcode=a1f51d94743c2eb38fd120b958ffed61
Text 4	https://www.uni-muenster.de/C-Test/demo/index.php

Spanish

Text 1	https://www.uni-muenster.de/C-Test/demo/index.php
Text 2	https://www.uni-muenster.de/C-Test/demo/index.php
Text 3	https://spraz-anmeldung.uni-hohenheim.de/cgi/ctest2.cgi?testcode=a42b0b5d23cb453b1a727798db973051
Text 4	https://buchsys.zfbk.uni-giessen.de/cgi/ctest2.cgi?testcode=482633343b9405607d1df53db6e3e145

Italian

Text 1	https://www.uni-muenster.de/C-Test/demo/index.php
Text 2	https://buchsys.zfbk.uni-giessen.de/cgi/ctest2.cgi?testcode=b1668daabe85f568ba4fd3c704ac7386
Text 3	https://spraz.uni-hohenheim.de/fileadmin/einrichtungen/spraz/docs/C-Test_Loesung_Italienisch.pdf
Text 4	https://buchsys.zfbk.uni-giessen.de/cgi/ctest2.cgi?testcode=b1668daabe85f568ba4fd3c704ac7386

Appendix M: Language Instrument Pilot Statistics
German

Text	Native Speakers $n = 5$	Intermediate (B1) $n = 5$	Advanced (C1) $n = 5$	Overall Comparison	NS vs Intermediate	NS vs Advanced	Intermediate vs Advanced
Jugendherberge	92.8% (M = 23.2; SD = 1.64316767)	53.6% (M = 13.4; SD = 4.03732585)	71.0% (M = 17.8; SD = 2.77488739)	F (2, 12) = 13.536, $p = <.001$	($p = <.001$)	($p = .044$)	($p = .106$)
Lernen	93.6% (M = 23.4; SD = 0.54772256)	47.20% (M = 11.8; SD = 3.1144823)	70.0% (M = 17.6; SD = 3.78153408)	Welch's F (2, 5.99) = 34.612, $p = <.001$	($p = .002$)	($p = .055$)	($p = .070$)
Nahrungsmittel	91.2% (M = 22.8; SD = 1.09544512)	51.20% (M = 12.8; SD = 2.16794834)	73.0% (M = 18.2; SD = 3.27108545)	F (2, 12) = 22.639, $p = <.001$	($p = <.001$)	($p = .030$)	($p = .012$)
Benzinpreise*	84.8% (M = 21.2; SD = 0.83666003)	41.6% (M = 10.4; SD = 1.51657509)	71.0% (M = 17.8; SD = 2.58843582)	Welch's F (2, 6.855) = 88.829, $p = <.001$	($p = <.001$)	($p = .086$)	($p = .003$)
Vereine	99.2% (M = 24.8; SD = 0.4472136)	48.0% (M = 12.0; SD = 1.58113883)	80.0% (M = 20.0; SD = 2.12132034)	F (2, 12) = 87.111, $p = <.001$	($p = <.001$)	($p = .001$)	($p = <.001$)
Hiddensee*	91.2% (M = 22.8; SD = 2.28035085)	46.4% (M = 11.6; SD = 0.89442719)	75.0% (M = 18.8; SD = 2.16794834)	F (2, 12) = 45.159, $p = <.001$	($p = <.001$)	($p = .019$)	($p = <.001$)
Dunkelheit*	97.6% (M = 24.4; SD = 0.54772256)	40.8% (M = 10.2; SD = 1.09544512)	78.0% (M = 19.4; SD = 1.94935887)	F (2, 12) = 146.830, $p = <.001$	($p = <.001$)	($p = .008$)	($p = <.001$)
Vitamine*	92.0% (M = 23.0; SD = 1.22474487)	43.2% (M = 10.8; SD = 0.4472136)	86.0% (M = 21.4; SD = 1.81659021)	F (2, 12) = 131.880, $p = <.001$	($p = <.001$)	($p = .189$)	($p = <.001$)
Average	92.8% (M = 23.2; SD = 0.01955761)	46.5% (M = 11.6; SD = 0.04703722)	75.5% (M = 18.8; SD = 0.08314145)	F (2, 21) = 193.958, $p = <.001$	($p = <.001$)	($p = <.001$)	($p = <.001$)

French

Text	Native Speakers $n = 5$	Intermediate (B1) $n = 5$	Advanced (C1) $n = 5$	Overall Comparison	NS vs Intermediate	NS vs Advanced	Intermediate vs Advanced
Régions*	96.8% (M = 24.2; SD = 1.30384048)	52.0% (M = 13.0; SD = 1.58113883)	83.2% (M = 20.8; SD = 2.28035085)	F (2, 12) = 52.617, $p = <.001$	($p = <.001$)	($p = .033$)	($p = <.001$)
Chiens*	94.40% (M = 23.6; SD = 1.94935887)	48.8% (M = 12.2; SD = 0.83666003)	76.0% (M = 19.0; SD = 2.44948974)	F (2, 5.99) = 34.455, $p = <.001$	($p = <.001$)	($p = .013$)	($p = .002$)
Brel	91.2% (M = 22.8; SD = 1.30384048)	41.6% (M = 10.4; SD = 1.14017543)	69.6% (M = 17.4; SD = 3.91152144)	Welch's F (2, 7.279) = 117.761, $p = .015$	($p = <.001$)	($p = .073$)	($p = .031$)
Internet	80.8% (M = 20.2; SD = 1.92353841)	49.6% (M = 12.4; SD = 1.14017543)	63.2% (M = 15.8; SD = 4.2661458)	Welch's F (2, 6.849) = 27.945, $p = .001$	($p = <.001$)	($p = .175$)	($p = .293$)
Nutella	91.2% (M = 22.8; SD = 1.92353841)	45.6% (M = 11.4; SD = 1.14017543)	77.6% (M = 19.4; SD = 2.50998008)	F (2, 12) = 45.469, $p = <.001$	($p = <.001$)	($p = .052$)	($p = <.001$)
Les députés entre deux villes	98.4% (M = 24.6; SD = 0.54772256)	56.8% (M = 14.2; SD = 0.83666003)	85.6% (M = 21.4; SD = 1.67332005)	F (2, 12) = 112.000, $p = <.001$	($p = .003$)	($p = <.001$)	($p = <.001$)
Etudier en France*	93.6% (M = 23.4; SD = 0.89442719)	45.6% (M = 11.4; SD = 1.67332005)	76.8% (M = 19.2; SD = 0.83666003)	F (2, 12) = 129.349, $p = <.001$	($p = <.001$)	($p = <.001$)	($p = <.001$)
Le vélo*	92.0% (M = 23.0; SD = 1.41421356)	48.8% (M = 12.2; SD = 0.83666003)	80.8% (M = 20.2; SD = 0.83666003)	F (2, 12) = 138.588, $p = <.001$	($p = <.001$)	($p = .005$)	($p = <.001$)
Average	92.3% (M = 23.0; SD = 0.02138925)	48.6% (M = 12.2; SD = 0.01917029)	76.6% (M = 19.2; SD = 0.05972855)	F (2, 21) = 196.154, $p = <.001$	($p = <.001$)	($p = <.001$)	($p = <.001$)

Spanish

Text	Native Speakers n $\equiv J$	Intermediate (B1) $n = 5$	Advanced (C1) $n = J$	Overall Comparison	NS vs Intermediate	NS vs Advanced	Intermediate vs Advanced
Actividades*	85.6% (M = 21.4; SD = 3.50713558)	55.2% (M = 13.8; SD = 1.09544512)	72.8% (M = 18.2; SD = 2.16794834)	Welch's F (2, 6.721) = 15.177, $p = .003$	($p = .015$)	($p = .261$)	($p = .016$)
Televisión*	83.2% (M = 20.8; SD = 1.30384048)	59.2% (M = 14.8; SD = 2.16794834)	84.0% (M = 21.0; SD = 2.00000000)	F (2, 12) = 17.904, $p < .001$	($p = .001$)	($p = .986$)	($p = .001$)
Desempleo Demo	80.8% (M = 20.2; SD = 2.04939015)	60.8% (M = 15.2; SD = 1.92353841)	73.6% (M = 18.4; SD = 2.50998008)	F (2, 12) = 6.775, $p = .011$	($p = .012$)	($p = .449$)	($p = .107$)
Entrenamiento	87.2% (M = 21.8; SD = 1.30384048)	59.2% (M = 14.8; SD = 1.4832397)	80.0% (M = 20.0; SD = 2.12132034)	F (2, 12) = 23.595, $p < .001$	($p < .001$)	($p = .247$)	($p = .001$)
El español en EE.UU.*	98.4% (M = 24.6; SD = 0.54772256)	52.8% (M = 13.2; SD = 1.30384048)	79.2% (M = 19.8; SD = 1.09544512)	F (2, 12) = 153.563, $p < .001$	($p < .001$)	($p = .001$)	($p < .001$)
El estado de ánimo en el invierno	92.0% (M = 23.0; SD = 1.22474487)	54.4% (M = 13.6; SD = 0.54772256)	84.8% (M = 21.2; SD = 2.16794834)	F (2, 12) = 57.446, $p < .001$	($p < .001$)	($p = .196$)	($p < .001$)
Columna Conmemorativa (1968)*	89.6% (M = 22.4; SD = 0.89442719)	52.8% (M = 13.2; SD = 1.09544512)	76.0% (M = 19.0; SD = 1.22474487)	F (2, 12) = 92.743, $p < .001$	($p < .001$)	($p = .001$)	($p < .001$)
Cuevas prehistóricas de Asturias	90.4% (M = 22.6; SD = 1.34164079)	64.8% (M = 16.2; SD = 1.30384048)	84.4% (M = 21.2; SD = 1.92353841)	F (2, 12) = 23.583, $p < .001$	($p < .001$)	($p = .390$)	($p = .001$)
Average	88.4% (M = 22.1; SD = 0.02770379)	57.4% (M = 14.35; SD = 0.05164785)	79.4% (M = 19.85; SD = 0.06268971)	F (2, 21) = 83.767, $p < .001$	($p < .001$)	($p = .006$)	($p < .001$)

Italian

Text	Native Speakers n $\equiv J$	Intermediate (B1) $n = 2$	Advanced (C1) $n = 2$	Overall Comparison	NS vs Intermediate	NS vs Advanced	Intermediate vs Advanced
Cassaforte	89.6% (M = 22.4; SD = 1.81659021)	62.0% (M = 15.5; SD = 0.70710678)	68.0% (M = 17.0; SD = 2.82842712)	F (2, 6) = 11.931, $p = .008$	($p = .014$)	($p = .040$)	($p = .744$)
Ladri*	92.8% (M = 23.2; SD = 0.83666003)	62.0% (M = 15.5; SD = 2.12132034)	86.0% (M = 21.5; SD = 3.53553391)	Welch's F (2, 1.442) = 8.566, $p = .158$	($p = .007$)	($p = .567$)	($p = .045$)
Movenpick	88.8% (M = 22.2; SD = 1.30384048)	38.0% (M = 9.5; SD = 0.70710678)	68.0% (M = 17.0; SD = 4.24264069)	Welch's F (2, 2.149) = 104.508, $p = .007$	($p = .001$)	($p = .062$)	($p = .030$)
Hashish	92.0% (M = 23.0; SD = 1.58113883)	58.0% (M = 14.5; SD = 0.70710678)	62.0% (M = 15.5; SD = 2.12132034)	F (2, 6) = 28.644, $p = .001$	($p = .002$)	($p = .004$)	($p = .824$)
Lo scandalo Parmalat*	95.2% (M = 23.8; SD = 0.83666003)	56.0% (M = 14.0; SD = 0)	74.0% (M = 18.5; SD = 0.70710678)	F (2, 6) = 133.566, $p < .001$	($p < .001$)	($p < .001$)	($p = .003$)
Cani	90.4% (M = 22.6; SD = 0.54772256)	68.0% (M = 17.0; SD = 1.41421356)	80.0% (M = 20.0; SD = 1.41421356)	Welch's F (2, 1.486) = 11.914, $p = .122$	($p = .148$)	($p = .329$)	($p = .288$)
Ecologia*	92.8% (M = 23.2; SD = 0.4472136)	56.0% (M = 14.0; SD = 1.41421356)	70.0% (M = 17.5; SD = 0.70710678)	Welch's F (2, 1.576) = 63.405, $p < .031$	($p < .001$)	($p < .001$)	($p = .010$)
La cucina Toscana*	90.4% (M = 22.6; SD = 0.89442719)	50.0% (M = 12.5; SD = 2.12132034)	74.0% (M = 18.5; SD = 0.70710678)	Welch's F (2, 1.970) = 131.880, $p = .037$	($p = .118$)	($p = .027$)	($p = .200$)
Average	91.5% (M = 22.88; SD = 0.01118034)	56.3% (M = 12.28; SD = 0.01767767)	72.8% (M = 18.18; SD = 0.06717514)	F (2, 21) = 51.760, $p < .001$	($p < .001$)	($p < .001$)	($p < .001$)

Focus Group 1

Int: Thank you for your time today and your introductions.

Int: I would like to start with a couple of language questions and then go onto questions regarding your personal development.

Int: How do you feel your language ability has developed over the course of the year abroad?

P1: I think when I was in the situation when nobody spoke English, I think my fluidity and just my ability to stay something instead of worrying about how to formulate the grammar changed, so my confidence in speaking the language definitely improved, but as much as the actual accuracy of language, I think that slipped a little bit as I didn't have somebody telling me "oh you need to conjugate that right!"

P2: Yes, I would say that my listening improved the most, I think that was because studying, the classes were in Italian, so I was forced to listen for an hour and a half at a time to someone speaking Italian, so that definitely helped. I think speaking wise, potentially more confident, although I found that I was in a lot of situations that weren't with people who only spoke Italian – there was a lot of time there was a lot of English, and also doing ERASMUS, there was a lot of ERASMUS students who only spoke English as a form of communication and Italian wasn't really used. And with writing, apart from a few Italian lessons when I was studying, I didn't do as much writing as I would have wanted to and so I don't believe my writing developed as much as I wanted it to.

Int: Have you perceived any frustrations with your language development?

P1: I found it really frustrating in Austria because I was living in quite a touristy area and as such, they were able to speak English and also because they speak a dialect of German, when I try to speak German with them, they wouldn't understand or just realise that it was easier to speak in English. It was really frustrating because I was there to speak German and so I would say that my German didn't really improve that much.

P2: My frustrations as that I continued to find myself in settings where there was a lot of English. I thought that I would be more immersed in purely Italian settings and as much as I was in some respects, I wasn't using my Italian as much as I wanted to be. I think as well, doing it at post A-level, I hit a barrier where I was struggling to really hit the next level of proficiency. I feel I didn't have the opportunities to use complex grammar structures for example, so I feel like I didn't hit the level I really wanted to be at because I continually found myself in situations where complex language wasn't required. Language was continually being spoken in quite chilled and informal settings.

Int: Do you feel you had matched expectations regarding your linguistic development?

P2: In talks with members of staff, they really sold the year abroad as an immersion experience and if you feel isolated, its ok, because you are going to be immersed. So I had that expectation going out to Italy and there's an expectation on return home that you experienced that. I really didn't get that experience which was quite frustrating to really communicate to members of staff. Your development and immersion really aren't as stark as you would think. It might be little things, like I have noticed my listening in class has really improved, which is a lovely surprise, but it wasn't such a stark improvement where I thought ... wow! I'm incredible now.

P2: Yes, definitely, I think when I was preparing myself to go out, I was preparing myself to fully immerse myself at every opportunity and I remember about halfway through my time in France I just felt totally exhausted ...

P1: Oh yeah, definitely

P2: ... with just trying to communicate in French because you have to put in so much effort to seek out the opportunities, instead of just staying in and watching something in English. Looking back now, it is slightly frustrating that I did not put more effort in, but I do think I have become more confident in my language ability, but I wouldn't say that I am like a native speaker level!

P1: I think also, it is the expectation of like, that you are going out there for a purely language immersive experience, when in reality when you are going somewhere for months on end, that is your new home and so there is a sense that you are going to a new place but that you have to live a different life because there is belief that you have to immerse yourself in it constantly and this is a lot of effort both physically and mentally. In reality, I just wanted to go home and watch Netflix because that is what I would do in my own home. So there was this contrast between the expectation of it being like a year abroad or holiday, but then the contrast that I am living here, I can't afford for it to be like a holiday.

Int: Upon reflection, how do you think your personal development has changed by going abroad?

P3: I think overall, I feel more confident being by myself, but also with other people, and more comfortable in both of those situations. I also now find it easier to meet and talk with new people. Over the year, however, I felt there was a point where I was tired of having to meet new people, and tired of constantly feeling like I needed to go out and see things. I felt at certain points, I just felt exhausted and needed a break from it all. After taken a short break, I would then pick up my want to go out and explore again. I think overall, the main thing that has changed for me is confidence in a number of areas.

P1: I would say that the balance between curiosity and being comfortable by yourself is huge. I have definitely noticed that I now seek out new opportunities and much more open to going to new societies etc. And inviting people over, if we have only met them once because when you are abroad that is how you make friends! But I am also much more aware now of when I have had enough; I have had enough and when I want to go to my room and chill out, I am comfortable of saying so. I am much more aware of being more extraverted and more introverted and I wasn't so clear of this distinction before going abroad. I am much more aware that I need to balance these two behaviours.

P2: I think definitely the ability to decide spontaneously on your own that you are going to do this, or you are going to do that because when on a year abroad you have no one telling you what to do and to just get out there and be confident and not afraid to try new things. I also have now developed the ability to say no to things. You understand yourself better and the things that you are able to cope with because you have faced so many new things which you haven't faced before, coming back to something so familiar you look at it from a different perspective. I have found my attitude to be much more like, ok so what do I need from this year? Instead of not going out and feeling I'm missing out, I'll stay in and not be so worried because I know that this is what I need.

P1: I think what has caused me frustration has been that I found myself to be a little less tolerant of other people. Hearing people complain, for example, that they don't have anyone they know

in their lecture and that they will be lonely is really frustrating because that isn't really a problem after facing what I have faced on a year abroad.

P2: Yes definitely! It is the contrast between people who did go on a year abroad, and those who didn't. It is like those who haven't been on the year abroad show similar behaviours to what we were showing in our second year.

P1: I would say I am more resilient to negative events but also more conscious of the effect negative events have on me and the ways to potentially combat that. I think the things that people find upsetting are now less worthy in my eyes of actually being upset over compared to what I experienced on my year abroad.

P3: Totally agree! My definition of a negative event has changed. For example, sitting on my own in a lecture wouldn't worry me now or be considered as a problem. I think I also view less stuff as a negative event.

P1: I think I'm a little dismissive of people's problems because the threshold of what is a problem has now changed.

Int: Would you say you have become less worried about things after your year abroad?

P2: I feel so much more in control in Britain because it is the familiar. The language is English, and I can ask people for help. My parents are in the south and this doesn't feel far away anymore. I just feel I'm much more in control of my life.

P1: I feel a lot more focused having come back. My year abroad didn't really have a focus. I studied but I didn't have exams. I worked but didn't have a lot to do so I had no focus. Upon returning, I have a lot more focus and appreciate being busy and my time-management is much better.

P3: I feel less worried about the future, I have done something out of my comfort zone, and I enjoyed it. I like having the structure of university, and I am looking forward to getting a job but I'm not as worried about all this as much before. I feel a lot less stressed about this year than all my friends who were in their third-year last year.

P2: I think it is because your outlook has now changed. You'll have bad times but you'll get over them and life continues. You will still find something enjoyable.

Int: Were there any events that triggered personal development?

P2: I think one thing which brought me anxiety was my accommodation. I was living in a boarding house where I wasn't allowed any visitors and had no kitchen. This constriction and not having my own independence was difficult and definitely increased my anxiety and limited my curiosity. However, when I moved, this great sense of independence came over me and all of a sudden, I found I was so much more curious and appreciated freedom much more.

P1: I think for me it was negative events which triggered personal development. I had accommodation difficulties and it was these situations which allowed development. It is all about contextualising the issue. For example, thinking where am I going to live in two days and all this makes you much more appreciative for what you have here in Britain.

P2: It is all about looking back and you think, if I managed to overcome the event and survive it what else can I achieve and overcome. It is all your own threshold changing because it these events which now serve as your baseline.

P3: I also had accommodation difficulties, but I turned it into a positive. The accommodation itself was a shared room and this meant that I went out and explored. I rarely spent a weekend in Prague, I travelled and during the days, even if not that busy, I went out to coffee shops etc. Just to get out the house.

P1: Yes, it really is the experience of negative events which change you the most. I found my job really difficult, I really disliked it, but ultimately my Dad made me see sense. You know, he said, working in a job which you find difficult or uncomfortable is the real world. He said that although you may not enjoy it, you will take something out of this experience and having this experience is really important. You will know how to deal with a difficult boss etc. It is trying to take the positives out of the negative events and knowing you will develop from them. I know a lot of people who avoid going into difficult situations for fear of failure or for feeling disappointment. Ultimately, you will grow from these experiences and going abroad have made me appreciate this so much.

P2: I felt that whenever anyone came to visit me (e.g., parents, friends), I went from really high to really low. You have the anticipation of showing someone your new life and then when they left, I felt really low.

P3: It was taking people to the airport and then turning around and getting the bus back on your own.

Int: Do you feel your expectations regarding personal development have they been matched?

P2: I don't think I took the time to understand what I was going to get out of the year. I just thought, well I will have a break from uni, improve my language and now I'm back and looking at jobs. I didn't really foresee the skills I would gain. Now I am back, and looking at jobs, there are so many new life skills I have develop which are valued. Everyone thinks the year abroad is going to be a really nice experience and I don't think that expectation is really realised. But I do think that there are a lot of advantages to going abroad that you don't really think about.

P1: Yes, I think the way the year abroad was presented makes you focus on your language. It was very much a language and culture experience, but in reality, it wasn't those things which I took positives or negatives from. It was more the personal development aspects but again, these weren't the things which I thought was going to develop and I didn't think I would feel so different as I do now than when I went, and I thought going back to university would just be the same as 2nd year, but I haven't found that at all. I think with regards to language and culture expectations, they really weren't matched, and I felt really disappointed and let down by certain aspects of this.

P3: I feel the opposite, I am not a language student and the year abroad wasn't something I had to do – it was a choice, which I think is quite a big difference. The focus wasn't on language, it was on everything else. I hadn't really thought about personal development beforehand, but I think it met my expectations and culturally it did too. It was lovely to have a break from University, I met new friends, went to new places and all these expectations were pretty much matched, although at different levels. On the whole, it met my expectations. For me, because it was a choice, in 2nd year, everyone was signing houses, and I hadn't applied yet, so I had to make the choice and be like ok, go ahead and look for a house, I am going to apply for a year abroad. At this point, although it was likely that I was going to get it, it wasn't certain that I would get it,

so the decision to do it at the early stage of 2nd year was much bigger I feel because for many I know, the thought of returning to the home university and everyone having graduated was depressing. Today, I don't recognise anyone in lectures! But because it was an active choice to go on a year abroad, I had taken much more thought into the advantages of disadvantages of going and decided it was worth going for these reasons and these reasons weren't big enough reasons for not doing it and then the year abroad did live up to these expectations.

Int: Regarding your well-being, how did this fluctuate over the year abroad?

P2: At the start, there was a little bit of anxiety because you don't know what is going to happen. But at the start you feel like you are on holiday, you get on the plane and then it starts to hit you. I think around halfway through my first country, I hit a real low – I was missing everyone at university, there were no English speakers around me and then my parents came out and it was a bit better. I think though that because I had such a difficult time socially in my first country, when I went to my second country, I suddenly picked up and I was able to socialise much more.

P1: I think I personally struggled on my year abroad, but I think this is because of pre-existing events before the year abroad which took place. Aside from that I think my accommodation difficulties really knocked me for my first placement. In hindsight, I look back and think I actually did have a really good time, but I think I just struggled with the personal and accommodation difficulties. But overall, I came out with some really lovely friends and a really fun experience. I think in my second question, because I didn't like my job I really struggled. I didn't meet a lot of friends and felt a lot of isolation. I struggled with the contrast between people seeing a year abroad as a long holiday and you are going to have a great time and then feeling like I wasn't and consequently the isolation caused by that, because I had people at home telling me it would be ok, I had a friend in Italy having a great time and who was telling me how wonderful it is and constantly telling me I should be having a great time which made it worse. I just felt like I should have been having a great time constantly, but I wasn't, and this made things more difficult to handle.

P3: I found the first three months, apart from the very initial anxiety, went really fast. Interestingly, I didn't get upset when two school friends left as I wouldn't have seen them anyway and I found the first three months as really positive for my well-being. However, the second semester was so much harder. All the friends I had met in the first semester were now leaving and this was more of a low point. There was a fallout from Christmas, it was cold, and most people were either studying for exams or were leaving. As such I returned home in February because I didn't want to spend a period of weeks abroad on my own. I enjoyed the second semester, but it wasn't as good. It was the same experience, but less ERASMUS students and fewer new people came. I didn't really meet any new people in the second semester and everyone's excitement to travel had lessened. There was less happening, less excitement to see new places and a lot more people from home came to visit which meant I was doing the same sort of trips every weekend with different people which was quite draining. I was also the last person to leave Prague, everyone had left, so was a little sad.

Int: How could the study have been improved?

P2: I think the trouble for me was having to have the internet on – if I was out and about, I didn't always have my data on. I knew I had to answer them, but I couldn't complete the daily questionnaires. When I was at work, it was also temperamental and also fitting the questionnaires into my own workday was difficult.

P1: I would say that the daily one was difficult to maintain. I forgot quite a few times and also felt that I hadn't done much in the day and didn't really know how to fill out the questionnaires.

P3: I agree with the others, but I would also say that completion (of state questionnaires) was dependent on where I was. If I was at home or in the library doing work and an email popped up I would do it, but if I was actually outside doing something and with other people, it was difficult to complete. As such, I felt my answers were not always representative of my experiences abroad because when I was out, I would miss the questionnaire but I'm not sure how this could be improved. Maybe, you could have one questionnaire at the end of the day but understand that retrospective answers may give slightly different answers.

P2: Also, some questionnaires came in on public holidays in one country and as such was travelling for a week. The answers I gave this week were not reflective of my general experiences of my normal week. It may not have been truly reflective of how I have been feeling.

P1: I liked the fact that the questionnaires made me reflective of how I feel, and I think it is the quizzes which have made me more aware of my mentality throughout the day – I feel stressed and why? What should I do to conquer this stress? It made me more aware of how have actually been feeling. It was nice to have the self-awareness and reminds you what it is like to feel normal. I feel better today which is a win compared to yesterday and that feels nice.

P3: I found the interaction questionnaire difficult as I found it difficult to visualise how long I have spent interacting in the language across a week. I tend to think more in terms of across the day. Moreover, I was only really able to give an approximation. I found it difficult to calculate how many hours I have spent using the language in hours across the week and as such sometimes felt I was guessing. It didn't really make sense as an answer I could come up easily. It should have been simple, but I found it wasn't.

P2: Sometimes I felt guilty because it made me feel that I haven't really used the language and I'm wasting my year abroad. When you put it in writing you it can be quite damning and makes you feel a little sad.

P1: I think also the English component was interesting as this always tended to be more (e.g., 4 hours of Netflix a day x7!) and again made me feel will am I getting that immersive experience.

Int: Would you do the year abroad again?

P2: I definitely would do one again, but I wouldn't want to go to the same place. But having that one year where you experience another language, culture and have all the experiences and challenges which go with that is great. It is a break from university, but you continue to learn, it just a different type of learning process. It has definitely improved me, and I would definitely do it again. Looking at careers though I now know that I don't want to work abroad.

P1: I would agree, and I would do it again because I like the personality development I have gone through. I feel more self-aware and feel more confident in putting myself out there. It has set me up better than if I hadn't have gone. In reality, I now know I definitely don't want to work in the country I was in – it is perhaps counterintuitive, but it has been helpful as I wouldn't want to leave all friends and family behind and then work out it isn't where I want to be.

P3: I would do it again and feel the same as other people. I considered the outside of Europe programmes but just applied for ERASMUS. I think I would apply again to the further of field ones if I did it again. I wouldn't want to live in Prague again. Throughout the entire nine months I met no one Czech Everyone I met were ERASMUS students and moreover the language barrier meant I wouldn't want to live there permanently. But in terms of your original question – would do it again!

Focus Group 2

Int: Upon reflection, how do you perceive your linguistic competencies to have developed?

P4: I think it has developed them enormously, I think for me, especially for me with my Russian, it was just a matter of confidence and I think the interpersonal skills you gain then reflect on your language speaking. In many more situations I wouldn't be scared to make a mistake in front of a native speaker because I now know that that is the best way to learn.

P5: I would agree with that. I feel like my Spanish definitely got better across the year abroad and likewise it was about the confidence. I find it more difficult to speak the L2 with native speakers, but abroad because I had to speak Spanish and there was no other alternative option it helped me build confidence and made speaking easier.

P6: I think as well that when you get to know people there who are native speakers and you trust them, you know them well, you don't feel as shy around them, you don't feel scared to speak their language because you know they aren't going to judge you.

P7: I'm not a language learner but I was sitting in on lectures in German for 8 hours a day, so I had no option but to learn fluently so that I understood and then I was examined, and my marks went to my degree.

Int: Which language component did you feel you benefited the most?

P7: Listening because I was listening most of the time.

P6: I would say speaking because I found that always the hardest and it was just nice to go over to the country and speak the native language.

P5: I would say speaking as well, and then probably listening.

P4: Yes, the same

Int: Do you perceive any frustrations in your linguistic growth?

P4: I feel like my expectations were matched but then when completing the ERASMUS online test at the start and end I hadn't change and there was the same distribution of scores but the components had change. For example, I had gone down in vocabulary which did not meet my expectations at all! So it was frustrating just knowing that I personally felt that I had this enormous growth and that not being reflected by one test on one evening.

P5: I would say that because there wasn't a high level of English spoken in my country, I found it easier to speak Spanish. It was frustrating at the start to try and get orientated to the immersion experience – from speaking Spanish a couple times a week to everyday was a change.

P6: I think for me it was just not being able to speak French as much as I wanted to. I was teaching in a school and I was teaching English so that was at least 12 hours a week where I couldn't practise French.

P7: Yes, the problem for me was that I was living with international students, so they all only spoke English really. They wanted to speak to me in English which meant I couldn't practice when I wasn't at the university.

Int: Upon reflection, how do you perceive your own personal development change?

P5: I think it is hard to say, I feel like, part of me wants to say yes, and part of me wants to say no. I feel that when I was on my year abroad, I think that I did change. There is such a big difference living away from your friends and family and a familiar environment for an extended period of time, so it had an impact on my confidence – it made me more confident to do things and also live independently because I had to do things on my own.

P8: I have greater perspective – when I'm getting stressed about a formative at least my lectures are in my L1 and not a foreign language. There are also times on the year abroad where you may be stressed or anxious, but you can't afford not to do something. For example, talking trying to get someone's help at a train station when neither of you speak the same language. It's not something I feel comfortable doing but the year abroad pushes you into those sorts of scenarios. When you return home, stressful or anxious situations don't appear so stressful or worrying. You have to fend for yourself.

P6: For me, I have become more open to change and not so stressed when placed in a new or unfamiliar environment. This year, I'm living with different people again and it is now a lot easier that I have gone to France and lived with a bunch of people I didn't know. It's a lot less stressful now at university getting to know new people.

P5: I think talkativeness. I find it is a lot easier to inject myself into a conversation. Last year I would have been a little more reticent about it. I totally agree with all the autonomy stuff and dealing with problems a lot more quickly. I definitely had this in Siberia, because if the WIFI wasn't working and you couldn't express yourself you just very quickly learn how to say it or explain your problem and resolve them a lot quicker. So maybe I'm less patient with myself and other things because I'm now just a bit quicker to deal with things.

P7: I wouldn't say I quite found that because Germany isn't too different from here and as such didn't really perceive any great changes.

Int: Were there any events which triggered personal development

P8: For me, because I was there for the whole year, most of my friends who I was close with in the 1st semester, including my housemates, all left and as such there was a really rough period between January and March where there was a really lonely transition between finding new friends but being in the same place. It wasn't the fact that I had changed country and therefore trying to find new friends. It was just as if they had all left me which was quite hard to process and learning how to cope with loneliness during this time was a big personality changer.

P6: Yes, I think for me it was near the end of the year when people started to leave, and we had also seen everything we wanted to see in the area. As such there was not much to do and I getting a bit bored and lonely.

P4: I dealt with sexual harassment which was a really negative event, but I actually feel really proud of myself and how I dealt with it. It was one of those things which you have to deal with really quickly and it also gave me more of a desire to really prove myself as a professional in the workplace so that was negative.

Int: Do you feel your expectations were matched regarding personal development?

P5: In terms of language I would say yes, but in terms of personal development, I feel like the year abroad promo can be a bit like an Instagram highlight reel and for me it wasn't like that.

Sometimes, when you aren't having the best time on your year abroad you are a bit like ... oh this is wrong; this doesn't match up what I have read about.

P8: I agree with you in terms people giving a really positive image of it. I think that often comes from the marketing team at a university. Whereas, when you speak to people who have already gone through the year abroad, would give a more a more expansive idea of what the experience was like including all the negative and positives of what the experience was like. I think a lot of blogs are visible and they do clearly say that not everything is great on the year abroad. I think they feel compelled to tell everyone to be realistic and an honest experience but sometimes universities do not do this.

P7: I just see it to be a year of your life. Some parts of it were amazing and some parts weren't so good. I don't think personally I would look back and say it was the best year of my life, but I also don't think I would say that it was the worst.

Int: Regarding your well-being, how did you find it to fluctuate?

P4: I think my well-being was similar to how it would have been at home. I had support from both my home institution and also in the country where I was living. Nothing really changed.

P6: I think at the start I was very busy, and I was getting use to everything and I had a lot of stuff to do. I would say however that as time was passing, I had done all the things I wanted too, and I wasn't meeting new people anymore. I feel that I was getting bored and with it depressed.

P8: I had a similar experience. I went home at Christmas, so the year was split. In the first half everything was new and exciting. We had a research project to do and I wasn't worried about this during the first half. When I came back to host country however my relationship ended, and I found my friends in the host country weren't as close to me as they were before Christmas. All of a sudden, a lot of things had changed and felt lonelier after Christmas.

P5: I found the first month of my second country difficult. It was January, issues at work, it was cold and dark and also living on my own. I then changed accommodation and it was lot better. The first month, I found it really hard to motivate myself and go out and do stuff and explore the city. I was just so exhausted that I was doing a 9 to 5 job for the first time, it felt daunting. But I think strangely for me because I was outside of at-home university, there wasn't that aspect of stress. In Siberia, I was at university, but we barely had any kind of homework stuff, just some Russian language stuff and the two essays I left until the end. So actually, I would say that my well-being was better abroad than at university because I didn't have that constant feeling that I should be doing work.

P6: I also found it less stressful abroad. Exams were split in January and July; it felt less work than if we had to wait to do all exams in the end like at university.

P7: It was lovely just to get out of the university treadmill. My friends were showing themselves working in the library and handing in essays and I felt like I was on a mini holiday.

Int: Were there any ways the study could be implemented differently?

P4: I think the difficult thing to log was the number of hours spent using the language. It was hard to remember and sometimes I felt I was guessing . Of course, if I wrote down my usage, answering that question would have been fine. Perhaps a scale like "a lot", 'not much time' may slightly be better.

P5: For me that was also the hardest part of the questionnaire. I did enjoy however, going through the questionnaires and it allowed me to consider my own emotions and behaviours. Perhaps I haven't spent enough time using the L2 or I have felt particularly anxious this month. Why is this and what can I do to try and change it?

P8: Yes, the only other thing I found difficult to fit into my time schedule was the multiple questionnaires in the day. I didn't for example, have my phone or laptop to hand at all time-point. I could have been out of WIFI range and it's kind of felt that once I missed a couple was there much more point or use in answering any more of the questionnaires.

Int: Would you do a year abroad again?

P8: I would say yes, I don't really have any desire to stay in the UK and now seriously considering further study or language teaching abroad.

P4: I agree, also a lot of people I know have said I don't want to stay in the UK. I don't think I would just do it for a year, I would want to stay there permanently. I really want to challenge myself again and love the process of integrating yourself into a new place and making it home.

P5: I would do it again but don't think I would do the same job.

P6: I would definitely do it again. I think if I could do it again, I would make more of an effort to get involved in some of the clubs that are on locally so that I can try and get new experiences as the year goes on and also meet a bigger range of people across the year.

Focus group 3

Int: How do you perceive your linguistic ability to develop over the year abroad?

P9: I went away for 14 months in two countries. What I have found coming back is the country that you were most recently in, your language skills are better in the speaking classes. Whereas if you do two languages, your second language is really suffering now so it can feel like one step forward, two steps back sometimes.

P10: I did Italian and French (Italian ab initio) and I did my Italian bit first, but I went backwards and forwards between countries. It was hard at first going from one to the other, but it has made my language skills more equal.

P11: As someone who doesn't study a language for a degree, but I was already reasonably good at the language I was learning, I found it extremely helpful. I was also in one country only so got a pretty concentrated period of development. I felt my language skills really improved in the second half of the year once I had picked up all the skills associated with conversation starter and got over the initial language hurdles.

Int: What aspects of language has developed over the year abroad?

P10: Speaking and listening in an informal setting and I have found that coming back I have forgot a lot of the grammar. I know the simple stuff; the stuff you use every day and phrases you don't really have to think about but having to think about more complex grammar stuff I have forgotten that, but I can say everyday language items such as slang proficiently.

P12: I think my Japanese has remained the same. In Japanese they quite often omit particles and are quite casual speakers, so I have now started to omit particles when I'm speaking.

P13: So I was in France and I wasn't a language learner. I had an ambition to learn some language but soon ended up dropping out and I found I learnt very little while abroad. I just learnt the very basic, enough to survive really.

P14: Yes, I also lived in France and I wasn't a language learner. I took some classes and found these helpful. I found people just switched to English when they tried communicating with me. I do think though that my listening improved more than anything because at the university I was at, most of the exchange students spoke amazingly good English so most social situations, unless you put yourself out there, it was difficult to pick up or improve in your listening skill.

Int: For those of you were language learners, did you feel you got that immersion experience?

P10: I deliberately worked and didn't study because I didn't want to be with other ERASMUS students. I also stayed in home stays and organised my year abroad so that I would be only with native speakers.

P12: With my programme, we could only study, there was no work option which means you spend a lot of time at the university where everybody speaks English. So I found that the best way to I improve my language was to leave the university and travel because the university wasn't very good at teaching its own language. I think it may depend on how you go about the year abroad (learning context). Maybe working would have been better because you are speaking to people on a regularly basis.

P9: I'm not sure if I entirely agree with the work aspect. In Spain, I got there and realised that I was the only person in the department who actually spoke Spanish. I went out there speaking Spanish pretty fluently when starting my year abroad – I had learnt it since I was a child, and then I found that I wasn't allowed to speak Spanish or Catalan in the office, so that was very difficult for me. But actually when things started to go wrong that was excellent because I was the only person who could deal with it and I was given a lot of responsibility. In France, I was a language assistant and so spoke English in the classroom. Having said that however, when you are in the staff room or meeting you are expected to use French and if you aren't doing it properly, well you have to, because that is your job. In Argentina, it was completely immersive but my boyfriend lives in Argentina with his family and so it was a natural immersive experience.

Int: Do you perceive any frustrations with your linguistic development?

P12: I think mine have because my lecture specified a standard of which she expected us to be at and I have the standard she expects.

P10: I wanted to be at C2 level and speaking with people who I was working with in each country said that I was definitely that level. But when I did the tests online because of my grammar and because I had forgotten it, I was at C1. It is frustrating because I was told one thing by native speakers but because I haven't done the sit-down revision, the test says I'm not.

P11: I had the same experience, even now when I speak German with native speakers, they say I'm excellent, but this isn't mirrored in the test. I think my German is very good now and it feels like a second native language, but it isn't technically at the level I wanted it to be at which is C2.

P9: I think I'm frustrated not in terms of the level because I set myself C1 as I know C2 is really living in that language for quite a while and so I sat my exams in C1 and passed them both so that wasn't really a problem. My frustration is that returning home with a dialectal language knowledge you can be perceived in a different way. You aren't necessarily perceived different by

the department but definitely by native speakers. If you speak a weird Spanish, you are perceived as not speaking real Spanish. I'm not sure if other people found that.

P10: We only speak Italian in Italy.

P12: Yes, different dialects can have an impact – this is definitely noticeable in Japan.

Int: Upon reflection how has living abroad developed your personal competencies?

P13: I don't think I have changed as a person; I think I got some new experiences, and I expected the years abroad to be better than I feel it was when I look back on it, but I think that was due to myself and the way I organised it. For example, I chose not to stay with a host family and lived in a student city where each nationality had their own building, so we were quite isolated there, but I didn't have the kind of development I was really expecting.

P14: I think I have definitely grown in self-confidence and putting myself in new situations because the year abroad is really a long line of different, new situations to what we are used to at our home institution. I also chose to live with a French family for most of the year. That also made different challenges to deal with, so I think I did change a bit over the year.

P11: I think I'm definitely more extraverted now and would score higher on Agreeableness. Over the year I felt myself becoming more attuned with other people's emotions and becoming more empathetic. I think it changed me a considerable amount, but this might not be due to explicitly the year abroad and more to do with simply my own personal growth.

P9: That's a good point and something that I didn't really think about when reflecting on my own development over the year. I think my change is through mixing with people who live in worst conditions to me through no fault of their own learn to live with those conditions and adapt. I now consider myself a diva and I have come back to university with the perspective that actually I'm really lucky and life isn't that difficult.

P10: Before coming to university, I worked and lived abroad and I think I changed as a person most then, which for a lot of people the year abroad is the first time they have been abroad. So I feel I have done a lot of the emotional changes already and felt prepared when going into the year abroad. But saying that I have become a lot more adaptable and I didn't think I had changed that much until getting back into university life and I now realise that I'm not the person I was in 2nd year. The reasons for this however, I can really put my finger on.

P12: I don't know because Japan is very similar to Britain with regards to their morals and outlook on life, so it quite similar but in a Japanese way and so I don't know if I have change.

Int: Were there any events which particular triggered change?

P10: I arrived in France not having anywhere to live and booked an Airbnb. I hoped to find accommodation within a 6-day period and I only found somewhere on the last day when I didn't have anywhere to stay that night, so that was quite stressful. But then I knew that if I could do that then I would be alright.

P9: I was really anxious when arriving in Barcelona. I arrived with nowhere to live and it was really expensive. I also arrived with little preparation, 5 days my final exams. My Airbnb was awful and dangerous. I had to leave, and I changed hostels four times and then I ended up living in a hostel for 3 months because accommodation was so expensive and hard to get. So this was hard built I accepted responsibility that I just couldn't ring the university and expect them to sort

it out. I also felt for the first time different as a woman and perhaps prejudiced against. You appreciate how safe Britain is.

P13: Nothing springs to mind. I wish I could say something.

P14: Likewise, I can't point to a specific instance where I felt some big situation changed my thoughts. It felt much more of a gradual process. I guess also because France is not hugely different from the UK and as such, I don't feel I had culture shock or that this change in culture impacted my behaviour.

P11: Well during my year abroad, my relationship broke up and also, I had several bad experiences on nights out. After a period of turbulence that changed my outlook on the purpose of the year abroad. When arriving I saw it as a chance for freedom and experimentation but after these experiences, I thought I had this chaotic experience and now it is time to slow down and I found out my own limits more.

Int: Were your expectations matched?

P10: I think you grow the most from experiencing negative events. You change and develop, learning from those negative experiences. It sounds airy fairy to say you grow so much as a person but what that actually means that there will be some really bad times which then mean you come out as a stronger person. So saying that it will be actually really beneficial for your growth means that you will actually go through some really low lows.

P9: I agree, and I think what I had seen at least was a very romanticised picture of Latin America having studied Latin American studies for a while and then went out I went out and saw poverty and poor living quality and we weren't given training as to how to confront that in the country. So when I got there, the first few weeks were rough. I couldn't go out alone for example and this is perhaps more for the last question but having a flat mate that had to flee her own country who was my age makes you realise how big your own problems are. It makes you appreciate what you have how small your own problems are in comparison to other people. So that is definitely personal development but on an academic level I don't really feel like I have developed in any way.

P13: A big difference for me was that all students from the UK came for a year whereas most of the other students are abroad for just half a year so the first half we would get to know people from other countries. But then it was like a new fresher week in January and it was hard to motivate yourself to find all new friends again. The second half I stayed more with my native country friends. I expected however, to go out more and socialise more with native speakers but I didn't find this at all. I basically socialised with just international students, although I admit that I could have done more, for example live in a home stay so that was a little disappointing.

P14: I did find that in our preparation, although only one workshop they were pretty realistic in the expectations they gave. I didn't find this in the study abroad fairs which did seem super optimistic. I felt in this workshop they communicated the idea that it is ok if it isn't the most amazing time, all the time.

P10: I think the language faculty sold it as an amazing experience, it will be perfect, and it will be the best year of your life. But because I had already been abroad, I knew that wasn't going to be the case and I knew there would be times when I was homesick. I also had a really strong support group from my church, and I spoke to fourth years who had already done who gave a much more realistic perspective. So I think that I was well prepared before going but I think if I did not have that experience and connections, I would have been more disappointed.

P9: And I think you are told that it is continual development, but I think when returning to university there was no talk beforehand about how difficult it is to readapt to a completely different environment such as studying when you haven't studied for the past year. It is just assumed that you will come back more able and adapt quickly but actually it can feel more like you are a fresher when you first come back – it is really strange!

P12: I do feel like the language department has some sort of halo around the year abroad. For example they have a photo competition which shows up all the best photos and best experiences.

P9: Another issue is that you are given all these expectations, for example, I was told that you could travel around South America, but you aren't told how you can finance them. So it's great to see the expectations of the year abroad, but you don't really know how to finance them.

Int: Looking in particular at your behavioural competencies in particular, how did this change.

P11: I think at the beginning I decreased in anxiety as I began to get comfortable with my surroundings. But then partly because outside factors and partly because of who I am, I became a lot more restless near the end of my year abroad and found myself increasing in anxiety. I think my curiosity increased, although I would say I am curious person anyway, but I find myself a lot more open to new experiences.

P14: I think in terms of anxiety, I think it decreased over the year and think it also made me more aware of my own anxiety when it is there. The first few weeks I was quite anxious as it felt like a fresher's week and then this re-emerges in January when most people we knew had left the university and it was a process of getting to know new people again. I was most aware of my anxiety when I didn't have a network in place. I would also say that I am calmer in unfamiliar and new situations.

P10: I think I became more curious and resilient and also more anxious which I wasn't expecting the latter, but I don't think that was to do with the year abroad.

P1: I probably became too relaxed in terms of work maybe – I didn't do too much work and university felt a distant thing – I got so relaxed that I forgot my visa had expired and that is when my anxiety all came to the surface – on the last day of my year abroad. I think my resiliency has improved because you have to deal with stuff on your own and you know what it is like to be alone. This is perhaps more if you are in a university because you have perhaps an expectation of being in a community and when I visited French university, this wasn't the case. You had to be resilient to keep making new friends.

Int: How do you feel your well-being fluctuated over the year abroad and what were the triggers of this?

P13: I think for me, the second half of my year, because all my friends left, I became more isolated in my room and I stayed in my room and didn't go out as much and do all the things I had expected to do. I think this triggered more anxiety because you are not doing what you think you should be going, for example integrating more. I also think I had this experience before because I'm not originally from Britain and I think it was more of a shock then because my language was not so strong. In French, I could speak English and that was cool, and I could communicate. I did not have the same skill set when first moving to Britain, so I felt the year abroad was easier because I had done it before.

P12: I found the teachers at my university to be condescending and rude and this really affected my anxiety and my general well-being. It was also limiting my language growth. So after about 8 weeks I decided that I needed to travel to get out of the university and also out of the city and that was a lot better. I also think a part of it was that teachers had expectations of you prior to going on a year abroad and they want you at this level when you come back, and you must also think about the assignment. The way they put it was that if you found yourself failing, if you reached out for them, it might be reflecting on you (as weakness). Perhaps, there needs to be more of an understanding by lectures before you left that if you were struggling that they wouldn't be judgemental about that. They advertise that you have well-being teams at your universities, but they don't really say how they will think about you. Like it is ok to fail and we aren't going to hold that against you.

P9: I think well-being comes with routine and I found changing countries meant it was very difficult to get a good welfare routine and I mean simple things like cooking dinner and getting enough sleep. My well-being was really dependent on the culture. I would say that my well-being was higher in South America than in Britain because it had much more of family culture and all of the things, I struggled with in England weren't apparent. But I definitely agree that there is pressure to not say you are struggling because you want to be that year abroad poster person at the end of the year abroad.

P11: I think culture is a good thing to mention in terms of well-being because in an academic environment, especially where I was, you would think if it was a high regarded university, there would be as much pressure as there is at university. In my experience however I found the opposite and people were independent and less connected with people around them, that independence came with a large amount of relaxation., No one was stressing out like they do at university for summative exams and so the culture and environment largely influences your well-being – this was definitely the case for me.

P10: I think over the year I understood what it was that was important for my well-being. I knew I needed to go outside and do exercise and I know that If I did those fundamental things needed for my well-being, I knew that wherever I went, I would be ok and that has also helped with re-entry.

P9: I think that it isn't good for your well-being in being the *foreigner*. I didn't have this in one country, but I did in other countries. I was the European and as such it was perceived that I needed to be looked after but sometimes this just makes you feel like a victim and that can make you feel worse, like you aren't capable of doing things. Maybe some people won't talk to you because they assume that you don't understand and sometimes you just want to show that I am ok.

* Due to a technical error, the final question did not record.

23.06.18

Participant Information Sheet

You are invited to take part in a research study entitled *'The traveller has many a tale to tell: Language development and personal change as a result of a year abroad'*. Please read this form carefully and ask any questions you may have before agreeing to be in the study.

The study is conducted by Mr Ian Moore as part of his PhD at Durham University.

This research project is primarily supervised by Dr Nadin Beckmann (nadin.beckmann@durham.ac.uk) from the School of Education at Durham University.

The purpose of this study is to capture data regarding how personality and linguistic ability can change in sojourners and non-sojourners. Those who do go abroad (sojourners) will serve as the intervention group, whilst non-sojourners will serve as the comparison group.

If you agree to be in this study, you will be asked to complete a series of questionnaires which are administered across the academic year.

You will first complete a questionnaire to capture demographic information (15 min). If you study a language, you will also be asked to complete a linguistic measure aimed at capturing your overall proficiency (20 min) for each studied language.

For every month throughout the academic year, you will be given a short (i.e., 5 minutes to complete) questionnaire. In addition, during one week per semester, you will be sent an even shorter questionnaire (i.e., 90 seconds to complete) that you will fill in four times a day. The study is conducted in this way to allow the researcher to establish how responses fluctuate during the day, across the week, and from month to month. Sometimes life simply gets in the way, so please do not worry if you miss a data collection point.

All questionnaires enable the researcher to learn more about your experiences both at-home and abroad, and how context (home vs abroad) has an impact on your behaviour. In total, participation in the study will take between two and a half and around four hours (for language students) across the entire academic year.

You are free to decide whether or not to participate. If you decide to participate, you are free to

withdraw at any time without any negative consequences for you.

All responses given, or other data collected will be kept confidential. The records of this study will be kept secure and private. All files containing any information you give are securely encrypted. All personal data will be held until September 2020 and destroyed after this date. You will be able to withdraw your data at any-time before this date by contacting the researcher. After this date, you will not be able to withdraw this data. This is because, after September 2020, all information held will be fully anonymous as email addresses will be destroyed.

Funding for this project comes from the Economic and Social Research Council (ESRC).

For your participation, you will receive your own personality profile throughout the year. This will display how your emotions/behaviours and situation perceptions changed across the year, together with how this related to changes in your linguistic ability. Each month you will also be entered into a prize with a chance to win one of five £10 vouchers to be spent in a number of online outlets (see: <https://www.voucherexpress.co.uk/content/e-vouchers-from-voucher-express.aspx>).

If you have any questions, requests or concerns regarding this research, please contact me via email at ian.j.moore@durham.ac.uk.

This study has been reviewed and approved by the School of Education Ethics Sub-Committee at Durham University (date of approval: 15/05/18)

Ian Moore



Leazes Road
Durham City, DH1 1TA

Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311

www.durham.ac.uk

Durham University is the trading name of the University of Durham
Durham University is the trading name of the University of Durham



Shaped by the past, creating the future

15/05/18

Ian Moore
ian.j.moore@durham.ac.uk

Dear Ian,

The traveller has many a tale to tell: Language development and personal change as a result of a year abroad: Study 2

Reference: 3096

I am pleased to inform you that your ethics application for the above research project has been approved by the School of Education Ethics Committee.

May we take this opportunity to wish you good luck with your research.

Yours sincerely,

A handwritten signature in black ink that reads "Nadin Beckmann".

Dr Nadin Beckmann
School of Education Ethics Committee Chair

Appendix Q: Participant Information Leaflet (Focus Groups)

1. Study title: The traveller has many a tale to tell: Language development and personal change as a result of a year abroad.

2. Background to the study: You are invited to participate in a research project that is designed to investigate your experiences whilst completing a year abroad. The study is being carried out by Ian Moore; a student at Durham University and being supervised by Dr Nadin Beckmann, who is an associate professor of Education in the School of Education at Durham University.

3. Why have I been invited? You have been invited because you completed a year abroad in the last academic year and are a student of MLAC at Durham.

4. Do I have to take part? Your participation is voluntary. Please contact the researcher if you have any questions regarding your involvement in the study. Even if you agree to participate now, you can withdraw at any time, without giving a reason, and without any consequences or any academic penalty. If you withdraw, your data will be removed from the study.

5. What will happen in the study? If you agree to participate, the study will involve a focus group discussion. This will comprise of all the participants getting together in a room and discussing their experiences abroad and your perceived changes. The focus groups will be audio-recorded and take a maximum of 90 minutes.

6. Are there any potential risks? If you elect to take part in the focus group, it may take up to a maximum of two hours of your time. You will be offered a break.

7. What happens to the research data provided? Any information collected during this research which can be identified with you will be treated with confidentiality. All data will be kept securely offline on both a computer hard drive and external hard drive of which will be password protected. The data files will also be encrypted. The files will not be kept on any open cloud drive such as Dropbox. The original data files will only be accessible to Ian Moore and his supervisor for transcription purposes only. If you wish to see any of your data, you can request this at any time. Data will be used for publication purposes, but these will contain NO personally-identifying information. In 2021, the files will be securely deleted.

8. Will the research be published? The research will be published in a thesis and hopefully, a subsequent publication. The researcher will ensure that the participant involves cannot be identified, e.g., by changing the names of the participant. You may request to be notified when the thesis is finished.

9. Who has reviewed this project? This project will be reviewed by the Ethical Committee of the School of Education at Durham University.

10. Who do I contact if I have a concern about the study or I wish to complain?

If you have a concern about any aspect of this project, please speak to the researcher (Ian Moore, ian.j.moore@durham.ac.uk) The researcher will acknowledge your concern and give you an indication of how he intends to deal with it. If you remain unhappy or wish to make a formal complaint, please contact the School of Education Ethical Committee at ed.ethics@durham.ac.uk

Appendix R: Focus Groups Consent Form

STUDY TITLE: The traveller has many a tale to tell: Language development and personal change as a result of a year abroad.

RESEARCHER DETAILS: Ian Moore, PhD Education, Durham University

PURPOSE of STUDY: You are invited to participate in a study investigating your experiences and perceived changes on a year abroad.

	Initial here
I have read the information sheet and had the opportunity to ask questions. If I asked questions, I have received satisfactory answers.	
I understand that my participation is voluntary, and I am free to withdraw at any time, without reason, and without any adverse consequences or any academic penalty.	
I understand audio recordings of the focus groups will only be made available to the researcher and his supervisor for transcription purposes.	
I understand that my data will be stored in password-protected electronic files, and the data deleted upon completion of the dissertation.	
I understand how to raise concerns with the researcher or make a complaint directly to the Research Ethics Committee at Durham University.	
I understand that excerpts from my interviews may be used in the thesis and possible publication.	
I understand that data from the project may appear in the researcher's thesis and subsequent possible publication, but the researcher will take steps to ensure this data will not be identifiable to me.	
I agree to take part in the study.	

Name of Participant:

Signature:

Date:

Name of researcher:

IAN MOORE

Signature:

Date:

Appendix S: Focus Group Ethical Approval

Dear Ian,

The following project has received ethical approval:

Project Title: *The traveller has many a tale to tell: Language development and personal change as a result of a year abroad;*

Start Date: *01 October 2016;*

End Date: *01 April 2020;*

Reference: *EDU-2019-06-27T19:33:21-srmw43*

Date of ethical approval: *19 July 2019.*

Please be aware that if you make any significant changes to the design, duration or delivery of your project, you should contact ed.ethics@durham.ac.uk for advice, as further consideration and approval may then be required.

If you have any queries regarding this approval or need anything further, please contact ed.ethics@durham.ac.uk

If you have any queries relating to the ethical review process, please contact your supervisor (where applicable) or departmental ethics representative in the first instance. If you have any queries relating to the online system, please contact research.policy@durham.ac.uk.