

Motivation and Multiglossia:
Exploring the Learning of
Arabic in UK Schools



Anna-Maria Ramezanzadeh

Department of Education
St. Hilda's College
University of Oxford

Thesis submitted for the degree of
Doctor of Philosophy

Hilary Term 2021

For my Father

ACKNOWLEDGEMENTS

بسم الله الرحمن الرحيم

The completion of this work would not have been possible without the help and support of countless individuals. I am indebted to my supervisors, Dr. Robert Woore and Dr. Lars-Erik Malmberg, for their guidance, patience, and unwavering support throughout this long journey. My gratitude, too, extends to Dr. Tamara Gray for precisely the same reason in a different realm. I am forever grateful to all the Arabic teachers and students who gave of their time and their selves during the course of this study. It is my hope that even if this thesis does not benefit them directly, it will help future generations of Arabic learners and teachers.

I have dedicated this work to my father, Sheikh Ramzy, to and for whom I am first and foremost thankful, primarily for his entirely unshakeable conviction, as far back as I can remember, that I would achieve this doctorate. My mother's trust in my ability to do this I also carried with me along this journey. My rock throughout this process has been my sister, Hannah, the most generous and consistently enthusiastic person I know. Without her friendship, support (even when she could easily have thrown my laptop out of a window), and constant supply of snacks, this thesis would not have been finished, at least not with my sanity intact. I am also indebted to my grandparents, particularly my grandfather, who many years ago sent me a newspaper clipping about this very subject, which sparked my initial interest in this field. I am immensely grateful also to Mae Zantout for her selfless generosity in reading through almost the entirety of this thesis in record time, and her compassion during my master's for which I still owe her my sincere thanks.

There are countless others – friends, family, and colleagues – whose help, kindness, and encouragement were of intrinsic value to this endeavour. My gratitude to them is no less present even if their names may be absent.

ABSTRACT

Motivation has been identified as a core mechanism in promoting achievement in language learning. Scant research, however, has been conducted into this construct in the field of Arabic as a Foreign Language (AFL). Arabic is multiglossic—consisting of a spectrum of standard, classical, and colloquial varieties—and has distinct ties to religion, thus presenting a unique case in terms of language learning motivation. This research examines the motivation of UK secondary school students studying for the General Certificate of Secondary Education (GCSE) in Arabic, using the lens of Dörnyei's (2009) Second Language Motivational Self System (L2MSS).

The central hypothesis of this study—that motivation is a multi-componential, dynamic, and language-variety specific construct, which interacts with proficiency—is investigated by means of a longitudinal, mixed-methods design. The study adopts elements of Complex Dynamic Systems Theory (CDST; Dörnyei & Ryan, 2015), approaching motivation as a system within which there are multiple constituents and mechanisms, each themselves complex, and which interact with each other, often in a non-linear fashion.

Data was collected from an initial sample of 265 students studying the Arabic GCSE in multiple schools around England at the beginning and end of one academic year (2017-2018). Motivation was measured in relation to three broad Arabic varieties: *Modern Standard Arabic*; *Classical Arabic*; and *Dialects*, operationalised according to the pragmatic language functions of each. To capture the complexity of learners' motivational self-concepts in relation to these language varieties, a new set of instruments measuring learners' *Current*, *Ideal*, and *Ought-to* selves was designed and tested for reliability and validity. The responses to these instruments were subsequently analysed to determine the structural validity of the L2MSS in the context of learning Arabic in UK schools. The nature, strengths, and interactions of learners' multiglossic self-concepts, along with their scores on two proficiency tests, were explored using t-tests and path model analysis. This facilitated the identification of intra-language variability in learners' self-concepts, interactions between their variety-specific current

and future selves, and the relationships between the self-concepts and proficiency outcome measures, from both longitudinal and cross-sectional perspectives.

Along with this quantitative data, interviews were conducted with 95 participants, which offer in-depth insights into the workings of the L2MSS and the legitimacy and efficacy of the model, as it currently stands, in capturing the diversity and complexity of motivation in this specific language learning context. The quantitative and qualitative datasets are analysed both individually and in concert, so as to integrate the findings and to present as comprehensive a picture of GCSE Arabic learners' motivations as possible.

The findings offer provisional validation for the underlying tripartite structure of the L2MSS in the context of AFL in secondary schools, as well as providing empirical support for the conceptualisation of this language as multiglossic in nature. However, the results of the analyses undertaken in this study also: (1) challenge the strength of traditional links between self-concepts and outcome measures; (2) highlight issues with and propose expansions to the constructs of the Current and Ideal selves; (3) underscore the problematic nature of the construct of the Ought-to Self; (4) propose the integration of a new self-concept component, the Past Self; and (5) question the role of the central underlying mechanism of the model—self-discrepancy—in promoting progress in language learning. These findings have wide-reaching implications for policy, pedagogy, and research in the realms of Arabic learning and as well as the study of motivation theory within the field of second language acquisition.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	4
ABSTRACT	5
TABLE OF CONTENTS	8
LIST OF APPENDICES	11
LIST OF TABLES	13
LIST OF FIGURES	15
LIST OF ABBREVIATIONS	17
1. INTRODUCTION	18
2. LITERATURE REVIEW	22
2.1. Motivation	22
2.1.1. A Brief History of Motivation Research	22
2.1.2. The Second Language Motivational Self System	23
2.1.3. Complex Dynamic Systems Theory	38
2.1.4. Summary	40
2.2. Arabic	41
2.2.1. The Arabic Language	41
2.2.2. What should be taught?	50
2.2.3. Teaching Arabic in the UK	52
2.2.4. Heritage Learners of Arabic	57
2.2.5. Arabic Motivation Studies	59
2.2.6. Summary	69
3. METHODOLOGY	71
3.1. Research Overview	71
3.1.1. Rationale for Adopting a Mixed-Methods Approach	71
3.1.2. Researcher Positionality	73
3.1.3. The Second Language Motivational Self System (L2MSS)	74
3.1.4. Proficiency Measures	76
3.2. Research Questions	78
3.3. Research Design	78
3.3.1. Recruitment & Sampling	80
3.3.2. Ethical Considerations	81

3.4. Methods	86
3.4.1. Instrument Development	86
3.4.2. Piloting	106
3.4.3. Recruitment	109
3.4.4. Data Collection Procedure	116
3.4.5. Data Transformation and Treatment	119
3.4.6. Data Analysis	120
 4. FINDINGS	 127
4.1. Structural Validity of the L2MSS and the Self-Rated Proficiency Scales	128
4.1.1. Second Language Motivational Self-System	129
4.1.2. Self-Rated Proficiency	145
4.1.3. Relationship between Proficiency Tests and Self-Rated Proficiency	150
4.1.4. Summary	151
 4.2. The Nature of Motivational Self-Concepts	 153
4.2.1. Quantitative Findings	154
4.2.2. Gender Differences	168
4.2.3. Heritage Background Differences	171
4.2.4. Qualitative Findings	175
4.2.5. 'Different Arabics'	176
4.2.6. Current Self	180
4.2.7. Ideal Self	195
4.2.8. Ought-to Self	215
4.2.9. Past Self	227
 4.3. Longitudinal Developments in Self-Concepts Over Time	 236
4.3.1. Preliminary Analyses	236
4.3.2. Comparison of Average-Score Means Over Time	237
4.3.3. Summary	240
 4.4. The Relationship between Self-Concepts and Proficiency	 242
4.4.1. Full Sample Analyses	242
4.4.2. Gender Differences	246
4.4.3. Heritage Background	249
4.4.4. Path Model Analyses	252
4.4.5. Summary	275
 4.5. The Role of Self Discrepancy	 276
4.5.1. Calculating Self-Discrepancy	276
4.5.2. Descriptive Analyses	277
4.5.3. Comparing Self-Discrepancy Scores	278
4.5.4. Longitudinal Developments in Self-Discrepancies	279
4.5.5. Path Model Analyses	280
4.5.6. Summary	285
 4.6. Summary of Findings	 286
 5. DISCUSSION	 288
5.1. Structural Validity of the L2MSS	288
5.1.1. Overview of Findings	288
5.1.2. Discussion of Emergent Themes	289
5.1.3. Summary	292

5.2.	The Nature of Motivational Self-Concepts	293
5.2.1.	Current Self	293
5.2.2.	Ideal Self	304
5.2.3.	Ought-to Self	312
5.2.4.	Past Self	322
5.2.5.	Summary	325
5.3.	Longitudinal Developments in Self-Concepts and Proficiency	326
5.3.1.	Self-concepts	327
5.3.2.	Proficiency Scores	329
5.3.3.	Self-Rated Proficiency Scores	329
5.3.4.	Summary	330
5.4.	The Relationship between Proficiency and Self-Concepts	331
5.4.1.	New Outcome Measures	332
5.4.2.	Linear or Reciprocal?	335
5.4.3.	School-Level Differences – The Role of the Learning Experience (L2LE)	337
5.4.4.	Summary	339
5.5.	The Role of Self-Discrepancy	339
5.5.1.	Overview of Findings	339
5.5.2.	The Optimal Gap Hypothesis	341
5.5.3.	Summary	341
5.6.	Summary: The Future Self of Motivation or the Future of Motivational Selves?	342
6.	CONCLUSION	344
6.1.1.	Key Findings	344
6.1.2.	Strengths of the Study	350
6.1.3.	Limitations	351
6.1.4.	Pedagogical Implications	354
6.1.5.	Areas for Future Research	355
REFERENCES		358
APPENDICES		373

LIST OF APPENDICES

A - LITERATURE REVIEW	373
A.1. Word Frequency Band Analysis	373
A.2. Arabic Motivation Studies Summary Table	375
B - METHODOLOGY	387
B.1. Population of Schools teaching GCSE Arabic	387
B.3. Proficiency Tests	404
B.6. Procedure for Categorising Heritage Language Learners	414
B.7. Consent Form & Information Sheets (Parents and Students)	417
B.9. Interview Analysis Coding Framework	425
C - FINDINGS	429
C.1. Missing Survey Data Statistics for Self-Concept Scales	429
C.2. Missing Survey Data Statistics for Self-Rated Proficiency Scales	430
C.3. Descriptive Statistics – Motivation Scales	431
C.4. Correlation Matrix – Self-Concept Scale Items	432
C.5. CFA Model Fit Indices for Language Varieties	436
C.6. CFA Model Fit Indices for Three-Factor Structure	437
C.7. CFA Three-Factor Structure Factor Loadings and Variance and Covariance Coefficients	438
C.8. Reliability Scores – Self-Concept Scale Items	442
C.9. Self-Rated Proficiency Items Descriptive Statistics	444
C.10. Self-Rated Proficiency Correlation Matrices	445
C.11. CFA Fit Indices – Self-Rated Proficiency	448
C.12. Reliability Scores – Self-Rated Proficiency Items	449
C.13. Descriptives Statistics – Proficiency Tests	450
C.14. Correlations between Proficiency Measures	455
C.15. Descriptive Statistics – Motivation Scale Average-Scores	456
C.16. Pairwise T-Test Results – Comparison of Self-Concept Average-Scores – Between Varieties	457
C.17. Self-Concept Average-Score Descriptive Statistics – Gender Differences	459
C.18. T-Test Results – Between Group Comparisons of Self-Concept scores – Gender Background	461
C.19. T-Test Results – Gender Split Comparison of Motivation Scale Average-Score Means	462
C.20. Self-Concept Descriptive Statistics – Heritage Background Differences	464
C.21. T-Test Results – Between Group Comparisons of Self-Concept scores – Heritage Background	466
C.22. T-Test Results – Heritage Split Comparison of Motivation Scale Average-Score Means	467
C.24. T-Test Results – Longitudinal Changes in Self-Concept scores – Gender	471
C.25. T-Test Results – Longitudinal Changes in Self-Concept Scores – Heritage Backgrounds	472
C.26. T-Test Results – Proficiency Tests Changes Over Time	473
C.27. Mean Comparison – Self-Rated Proficiency Over Time	474
C.28. Proficiency Test Descriptives – Gender	475
C.29. T-Test Results – Proficiency Test Score Comparisons between Genders	476
C.30. T-Test Results – Proficiency Test Score between time comparisons within Genders	477
C.31. Self-Rated Proficiency Descriptive Statistics - Gender	478
C.32. Mean Comparison Results – Self-Rated Proficiency Comparisons by Gender	480
C.33. Mean Comparison Results – Self-Rated Proficiency between time comparisons within Gender	481
C.34. Proficiency Test Descriptive Statistics – Heritage Background	482
C.35. T-Test Results – Proficiency Test Score Comparisons between Heritage Backgrounds	483
C.36. T-Test Results – Proficiency Test Score between time comparisons within Heritage Backgrounds	484
C.37. Self-Rated Proficiency Descriptive Statistics - Heritage Background	485
C.38. Mean Comparison Results – Self-Rated Proficiency Comparisons by Heritage Backgrounds	486
C.39. Mean Comparison Results – Self-Rated Proficiency between time comparisons within Heritage Backgrounds	487
C.40. Frequencies of References to Node Categories by School	488

C.41. Notes on Consistency	489
C.42. Caterpillar Plots for Self-Concept and Proficiency Test Variables	494
C.43. Descriptive Statistics for Self-Discrepancy Sum Scores	499
C.44. Correlation Matrices for IS Plausibility and Self-Discrepancy	500
C.45. Self-Discrepancy Path Model Output	501

LIST OF TABLES

Table 2.1. Conjugation of 'to study' in the imperfect tense _____	43
Table 2.2. Heterophonic-homographs from root عقد _____	43
Table 2.3. Arabic Varieties (adapted from Wahba, 2006) _____	47
Table 2.4. Grade Distributions for GCSE Arabic—Summer 2020 _____	54
Table 2.5. Orientations of Students towards Learning Arabic (Ramezanzadeh, 2015) _____	62
Table 3.1. Motivation Survey Components _____	87
Table 3.2. Motivational Self-Concept Scale Opening Statements _____	89
Table 3.3. Motivational Self-concept Scale Components _____	89
Table 3.4. Language Variety-Related Skills included in Self-Rated Proficiency Scale _____	96
Table 3.5. Cloze Test Text Sources _____	102
Table 3.6. Overview of Schools and Sample _____	112
Table 3.7. Gender Distribution of Sample at T1 and T2 Surveys _____	113
Table 3.8. Home Languages of Participants _____	115
Table 3.9. Number of Years of Arabic Study prior to Year 10 _____	116
Table 3.10. Summary of Data Analysis Methods _____	121
Table 3.11. Phase 1 Descriptive Codes _____	123
Table 4.1: Motivation Scales—Missing Data Summary _____	132
Table 4.2: CFA Factor Loadings, Errors, and Item Covariances T1 _____	137
Table 4.3: CFA Factor Loadings, Errors, and Item Covariances T2 _____	138
Table 4.4: Motivation Scale Reliability Scores (T1 & T2) _____	144
Table 4.5: CFA Factor Loadings—Self-Rated Proficiency T1&2 _____	149
Table 4.6. Frequency of Cases and References to Varieties _____	162
Table 4.7. Cross-tabulation of variety-self references _____	162
Table 4.8. Paired Samples T-Test results—Comparison of Feelings towards Arabic & GCSE _____	164
Table 4.9. Language Variety Hierarchies in Male and Female Subgroups _____	169
Table 4.10. Language Variety Hierarchies in HL and NHL Subgroups _____	174
Table 4.11. Frequency of Arabic Future Self References _____	213
Table 4.12. Significant correlations between Arabic at school and Self-concepts _____	228
Table 4.13. References to AFS codes categorised by Past Experience _____	231
Table 4.14. References to Feelings codes categorised by Past Experience _____	232
Table 4.15. Mean Comparisons for Self-concepts by Past Experience _____	234
Table 4.16. Descriptive Statistics – Proficiency Scores T1 & T2 _____	243
Table 4.17. Independent T-Test results for comparison between HLs and NHL proficiency test scores at T1 & T2 _____	250
Table 4.18. Model Fit Indices for Language Variety Models and Overall Model – Self-Rated Proficiency _____	255
Table 4.19. R2 Estimates for Model 1 _____	258
Table 4.20. Model Fit Indices for Language Variety Models and Overall Model – Proficiency Tests _____	260
Table 4.21. R-Square scores for All Models _____	261
Table 4.22. Frequency of Grammar and Vocabulary Codes _____	265
Table 4.23. Cross-references to Grammar/ Vocabulary and Four Skills _____	266
Table 4.24. Intra-Class Correlation Coefficients for All Self-Concept and Proficiency Test Variables at T1 & T1 _____	271
Table 4.25. Classroom Means Diverging from Grand Mean per Self-Concept and Proficiency Test Variable _____	273
Table 4.26. Independent Samples T-Test—Self-Discrepancy Differences between Heritage Backgrounds _____	279

Table 4.27. Percentage of Variance Explained—Self-Discrepancy and Proficiency Measures

284

LIST OF FIGURES

Figure 2.1. The Interrelationship of the Motivational Variables and Criterion Measures	24
Figure 2.2. The Second Language Learning Motivation Self System	25
Figure 2.3. Hessel (2015)'s Motivation Survey Items	30
Figure 2.4: The Arabic Alphabet and Script	42
Figure 2.5. Infographic on Variance of Dialects in the Middle East	49
Figure 2.6. Schools offering Arabic in England and Wales	53
Figure 2.7. 'A taxonomy of human motivation'	65
Figure 3.1. The Second Language Motivational Self System	74
Figure 3.2. Arabic Multiglossic Motivation Matrix	75
Figure 3.3. Theoretical model of L2MSS construct interactions	77
Figure 3.4. Research Design and Sample Statistics	79
Figure 3.5. Sample of Motivation Scale Items (Modern Standard Arabic)	94
Figure 3.6. Y/N Test Items organised by Word Frequency Level	99
Figure 4.1. Summary of Research Questions	127
Figure 4.2. Missing Data Summary (Motivation Scales T1 & T2)	130
Figure 4.3. Missing Data Summary (Self Rated Proficiency Scales T1 & T2)	131
Figure 4.4. Three-Factor Structure CFA Diagrams	140
Figure 4.5. Bar graph comparing descriptive data of retained sample and attribution group	155
Figure 4.6. Comparison of mean scores per self-concept for each variety for the full sample	158
Figure 4.7. Bar graph comparing self-concept average-score means per variety at T1 & T2	161
Figure 4.8. Comparison of mean scores per self-concept for each variety for males and females	170
Figure 4.9. Comparison of mean scores per self-concept for each variety for HLs and NHLs	172
Figure 4.10. Bar graph for references to Emotions codes by gender	192
Figure 4.11. Bar graph for references to Emotions codes by heritage background	194
Figure 4.12. Bar graph for references to Arabic FS images	196
Figure 4.13. Bar graph for references to Arabic FS images by gender	208
Figure 4.14. Bar graph for references to Arabic FS images by heritage background	209
Figure 4.15. Bar graph for references to Ought-to Self codes by gender	225
Figure 4.16. Bar graph for references to Ought-to Self codes by heritage background	226
Figure 4.17. Bar graph for comparison between average-score means for self-concepts at T1 and T2.	238
Figure 4.18. Bar graph for comparison between average-score means for self-concepts at T1 and T2 for HL and NHL subgroups.	239
Figure 4.19. Bar graph for comparison of proficiency test scores at T1 and T2.	243
Figure 4.20. Bar graph for comparison of Self-Rated proficiency scores at T1 and T2.	245
Figure 4.21. Bar graph for comparison of Self-Rated Proficiency scores at T1 and T2 separated by gender.	248

Figure 4.22. Bar graph for comparison of proficiency test scores at T1 and T2 separated by heritage background. _____	250
Figure 4.23. Bar graph for comparison of Self-Rated proficiency scores at T1 and T2 separated by heritage background. _____	252
Figure 4.24. Factor Score Path Models for each language variety and Overall model _____	256
Figure 4.25. Factor Score Path Model with Proficiency Test variables – Overall Model ____	258
Figure 4.26. Factor Score Path Model with Proficiency Test variables – MSA Model _____	260
Figure 4.27. Factor Score Path Model with Proficiency Test variables – Dialect Model ____	262
Figure 4.28. Factor Score Path Model with Proficiency Test variables – Classical Model ____	264
Figure 4.29. Caterpillar Plot for Y/N Test scores per classroom _____	272
Figure 4.30. Hypothesised Self-Discrepancy Factor Score Path Model _____	277
Figure 4.31. Self-Discrepancy Factor Score Path Model – Overall _____	280
Figure 4.32. Hypothesised Optimum Discrepancy Plot _____	282
Figure 4.33. Scatter plot of Overall Self-Discrepancy at T1 against Cloze Test Scores at T2	283
Figure 4.34. Scatter plot of Overall Self-Discrepancy at T1 against Y/N Test Scores at T2__	283
 Figure 5.1. Future Self Trajectory _____	 321

LIST OF ABBREVIATIONS

AFL	Arabic as a foreign Language
AMTB	Attitudes/Motivation Test Battery
CFA	Confirmatory factor analysis
Cl	Classical Arabic
CS	Current Self
D	Dialect [Colloquial Arabic]
FS	Future Self
GCSE	General Certificate of Secondary Education
HL	Heritage Learner
IS	Ideal Self
KS	Key Stage
L1	First Language
L2	Second Language
L2LE	Second Language Learning Experience
L2MSS	Second Language Motivational Self System
LOTE	Languages Other Than English
MSA	Modern Standard Arabic
NHL	Non-Heritage Learner
NS	Native Speaker
OTS	Ought-to Self
PLE	Past Learning Experience
PPE	Past Personal Experience
PS	Past Self
S-EM	Socio-Educational Model
S-D	Self-Discrepancy
SR Proficiency	Self-Rated Proficiency
T1/T2	Time 1/Time 2
Y/N	Yes/No [Proficiency Test]

1. INTRODUCTION

Motivation has been identified as a cornerstone of learning; a powerful force which can promote success even in the absence of strong learning aptitude (Dörnyei, 2015). One of the most prominent theories in the field of second language (L2) learning motivation is the Second Language Motivational Self System (L2MSS). Originally formulated by Dörnyei and his colleague Csizér (2005), the model incorporates the psychological concepts of *possible selves* and *self-discrepancy theory*, through which motivation is conceptualised as a learner's desire to reduce the gap between their future and current L2 self-concepts (Dörnyei, 2015). The original tripartite model of the L2MSS combines the Ideal Self (IS), the Ought-to Self (OTS), and the L2 Learning Experience (L2LE) to operationalise and explain the dynamics behind language learners' motivations. In light of substantial challenges to this model, a revised theoretical framework and new instruments were developed for the purposes of this study, incorporating qualitative and quantitative measures of the IS and OTS, along with the Current Self (CS). The redeveloped model is applied in the context secondary school learners of Arabic as a foreign language in order to capture and explore their motivational self-concepts and the mechanisms at work behind them.

The Arabic language is unique in its sociolinguistic nature. It consists of a range of varieties, dialects, and registers, a phenomenon referred to in this study as 'multiglossia' (Badawi, 1973). The exact delineation and definition of the language's multiglossic components is the subject of continued debate (Omar, 2011). This study proposes a conceptualisation of the language consisting of formal (Modern Standard and Classical) and informal (Dialect) varieties, each of which is used in distinct but overlapping settings in Arabic-speaking communities (Wahba, 2006).

In recent decades, the demand for Arabic instruction in the UK has expanded, largely as a result of globalisation, shifting economic powers, and events in the Middle East which have reverberated around the world (Tinsley, 2015). One of the only standardised qualifications available for Arabic in the UK is the General Certificate of Secondary Education (GCSE). Despite the growth in its uptake, little research has been conducted into this qualification. An analysis of its utility, suitability, and accessibility are

becoming increasingly necessary as calls for its expansion are made (Tinsley & Board, 2013).

The integral nature of motivation in language learning, coupled with the rise in prominence of the Arabic language in the context of the UK, serves to highlight the absence of research on the Arabic learner population, and the dearth of theoretically based research into motivation amongst Arabic learners in general. These factors combined provided strong impetus for this study.

The following chapter provides an overview of the literature on the L2MSS, outlining its development from Gardner's (1985) Socio-Educational Model, and support for the Ideal Self construct in predicting intended effort outcome measures. Attention is then turned to presenting the challenges to the model and its application. Most prominent of these are: the inclusion of CS measures; the nature and function of the OTS; the role of the model in promoting objective measures of achievement; the application of the theory to languages other than English and in the context of younger learners; and the incorporation of qualitative and longitudinal design elements. The section ends with a discussion of Complex Dynamic Systems Theory (CDST) which has been proposed as a lens through which to capture and interpret the intricacies of the components and mechanisms at work within learners' L2 motivational profiles (Dörnyei & Ryan, 2015).

An outline of the Arabic language is next provided, detailing its multiglossic nature, and other challenging features entailed in its teaching and learning. Arabic is currently taught in a limited number of secondary schools in the UK, predominantly within the framework of the GCSE (Tinsley & Board, 2013). Details are provided of the structure of this course, and an examination is offered of the content and approach of the GCSE curriculum which, it is shown, features only the Modern Standard variety.

Tying the two strands of motivation research and the Arabic language together, the final section of Chapter 2 includes an overview of studies on the motivations of Arabic language learners. The limited number of L2MSS studies conducted in the context of Arabic learning are examined.

The Literature Review provides the basis and rationale for the focus of this study: the investigation of the nature and validity of the L2MSS in an under-explored context—secondary school learners of Arabic in the UK. The Methodology (Chapter 3) outlines the research design, which addresses the challenges arising from previous research by adopting a longitudinal, mixed-methods approach. The development of three new instruments are described: the Motivation Surveys, which incorporate measures of learners' self-concepts in response to the three Arabic varieties as well as measures of their self-rated proficiency; proficiency tests, which include Yes/No and cloze tests formulated using vocabulary and texts from the GCSE specification and sample materials; and the interview protocol designed to elicit in-depth information on learners' self-concepts and the role of other factors in the construction of their motivation. These instruments were designed and employed, and the results analysed, in response to five research questions:

1. What is the structural validity of:
 - a. the L2MSS construct as it relates to Arabic, as measured by the motivation survey scales?
 - b. the multiglossic self-rated proficiency construct, as measured by the self-rated proficiency scales?
2. What is the nature of Arabic learners' motivational self-concepts? How do learners' motivational self-concepts vary based on:
 - a. Arabic language variety?
 - b. Gender?
 - c. Heritage Learner status?
3. How do learners' motivational self-concepts develop over time?
4. What is the relationship between proficiency and self-concepts:
 - a. Within timepoints?
 - b. Between timepoints?
5. What role does self-discrepancy play in the L2MSS?

These research questions examine both the underlying components of the L2MSS—the motivational self-concepts—and the framework within which they operate and interact. Chapter 4 presents the findings in response to each of these research questions. The first establishes the structural validity of the L2MSS and self-rated proficiency measures, confirming the validity of a multiglossic conceptualisation of the Arabic language, and the three self-concepts. The second section provides an extensive investigation of the nature of these self-concepts in which findings from the qualitative and quantitative data are intertwined using the ‘Following the Thread’ approach, proposed by Moran-Ellis and colleagues (2006). Longitudinal changes to the self-concepts and proficiency measures are then explored, followed by an investigation of the interactions between these constructs. Finally, an exploration is undertaken of the self-discrepancy component of the L2MSS framework.

The Findings chapter is followed by an in-depth discussion (Chapter 5) of the results generated in response to the research questions, taking each in turn. Both the components of the L2MSS and the interactions between them are reviewed, with a CDST lens employed where appropriate. Important implications for both theory and practice stemming from these findings are highlighted throughout the course of the chapter. The study is concluded in Chapter 6 with a summary of its key findings and novel contributions to the field, an overview of its strengths and limitations, and implications for the study of L2 motivation and Arabic language pedagogy in the UK and beyond.

2. LITERATURE REVIEW

2.1. Motivation

The term ‘motivation’ is derived from the Latin *motus*, meaning ‘moving, or motion’, and the Old French term *motif*, meaning ‘will’, or ‘drive’ (Harper, 2010). In essence, motivation refers to ‘that which inwardly moves a person to behave a certain way’ (Random House, 2016), and in the social sciences it is used to refer to ‘the psychological foundation of any human behaviour or thought, subsuming all the factors that shape what we do or think’ (Dörnyei, 2020). The study of motivation thus attempts to identify reasons for individuals’ behaviour. In language learning, it is posited to be a reliable predictor of success, able to compensate for limitations in students’ aptitude and deficiencies in the learning environment (Dörnyei, 2015). Developing a framework for the measurement of a phenomenon which encompasses every factor impacting human behaviour, however, is no easy task. Whilst the term itself has ‘intuitive appeal’, and on the surface identifying motivated behaviour in ourselves and others seems a straightforward task, one of the most fundamental challenges in this field has been how to conceptualise and operationalise *motivation* (Dörnyei, 2020). This section offers a brief overview of the development of motivation theories, and explores recent developments in the field, focusing in particular on the most dominant framework employed to measure language learning-specific motivation: the Second Language Motivational Self System (L2MSS).

2.1.1. A Brief History of Motivation Research

In 1959, Gardner and Lambert likened the process of second language (L2) learning to that which a child experiences when learning their first language (L1), in which the child attempts to assimilate into their surroundings through a desire to be like members of their family and wider community (Gardner & Lambert, 1959). This desire to integrate into a community, which Gardner and Lambert termed *integrativeness*, formed the central tenet of their subsequent motivational model: the Socio-Educational Model (S-EM).

The S-EM incorporated a number of affective factors, all of which contribute to a person's drive to learn a language, measured using the Attitudes/Motivation Test Battery (AMTB) (Gardner, 1985). Whilst it is an oversimplification to reduce the model to merely two elements, the central concepts emerging from the theory were *instrumentality* and *integrativeness*. The latter Gardner (2001) defined as 'a general interest in learning the second language in order to come closer to the other language community' (p.5). Instrumentality, on the other hand, reflected 'more utilitarian values of linguistic achievement' (Gardner & Lambert, 1959, p.267), where learning is pursued for 'more practical concerns' (Gardner, Masgoret, Tennant et al., 2004, p.6).

While the S-EM has received much support over the past three decades, its application has become increasingly problematic. The concept of integration into an identifiable L2 community was applicable to the bilingual context of Canada in which Gardner and his colleagues first formed the model. However, applying it to other language contexts, in particular English, was difficult since a distinct group of L2 speakers to integrate into did not always exist.

2.1.2. The Second Language Motivational Self System

In the early 2000s, the language learning motivation field witnessed a reconceptualization of its cornerstone concept, integrativeness. In 2002, Dörnyei and Csizér ran a multivariate statistical analysis of motivational dispositions amongst Hungarian school students. They collected survey responses from 8,593 13 and 14-year-olds learning a mixture of five different L2s. Students were asked at two time-points to complete surveys investigating their motivation and two criterion measures: *intended effort* and *language choice*. The survey consisted of 37 items, from which seven factors emerged (Figure 2.1): *Integrativeness*, *Instrumentality*, *Vitality of the L2 Community*, *Attitudes towards L2 Speakers*, and *Cultural Interest*, *Linguistic Self-Confidence*, and *Milieu*.

Surprisingly, the latent factor, normally identified as integrativeness, the only factor mediating the criterion measures, had as its antecedents Attitudes towards L2 Speakers and Instrumentality (see Figure 2.1). The fact that integrativeness subsumed

Instrumentality, as well as learners' Attitudes towards L2 Speakers challenged the accepted notion of motivation as a desire to integrate, since it was also closely associated with more utilitarian orientations.

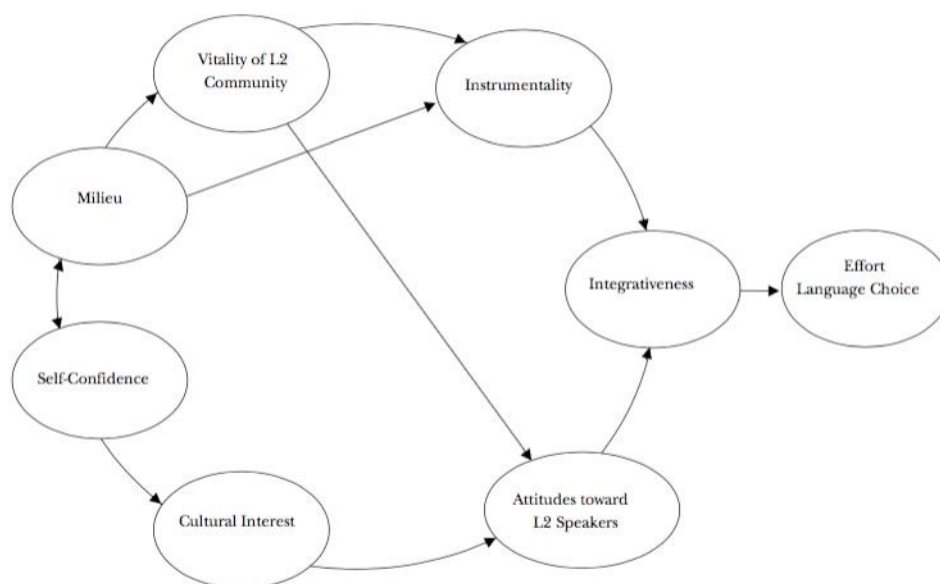


Figure 2.1. The Interrelationship of the Motivational Variables and Criterion Measures (Csizér & Dörnyei, 2005, p. 28)

These results led Dörnyei (2005) to offer a reinterpretation of this principal element of motivation. He identified the core of integrativeness as a psychological and emotional process associated with a person's identity. However, rather than viewing motivation through the paradigm of orientations towards the language community, Dörnyei reconceptualised motivation as an internally directed process within the framework of self-systems. In other words, whereas Gardner's definition involved the integration of oneself into an *external* locus, the *L2 community*, to Dörnyei, motivation represented the integration of the language into *one's own self-concept*, making the focus an *internal* one.

In formulating this revised model, Dörnyei drew upon two psychological theories: *possible selves* (Markus & Nurius, 1986) and *self-discrepancy theory* (Higgins, 1987). Possible selves theory purports that a person's future self-concept consists of images of who they would *like* to become, *might* become, and are *afraid* of becoming. Dörnyei integrates this with Higgins' theory self-discrepancy theory which argues that

motivation stems from the desire to reduce the uncomfortable gap between one's current and future self-concepts. The combination of possible selves theory, self-discrepancy theory, and the reconceptualization of integrativeness led to the development of the L2MSS; a tri-partite model consisting of the *Ideal L2 Self*, the *Ought-to L2 Self*, and the *L2 Learning Experience* (L2LE) (illustrated in Figure 2.2).

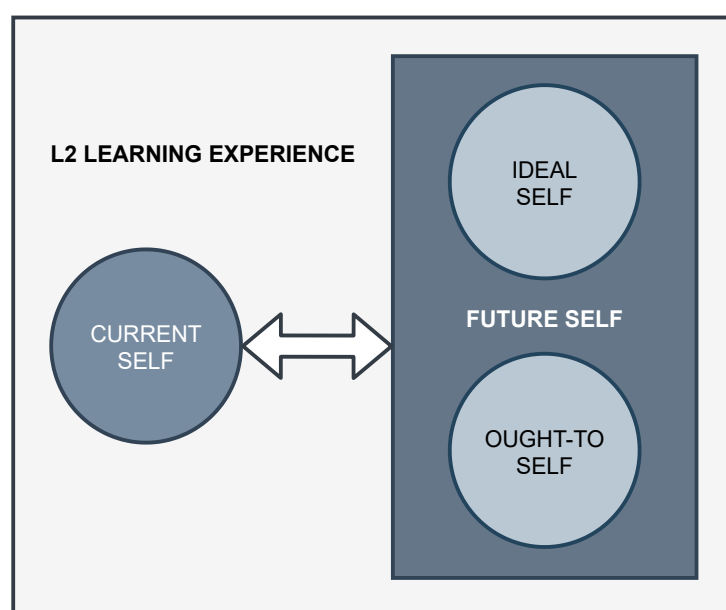


Figure 2.2. The Second Language Learning Motivation Self System

- *Ideal L2 Self (IS)*: refers to ‘the L2-specific facet of one’s ideal self’ (Dörnyei, 2009, p.29): ‘If the person we would like to become speaks an L2 [...] the IS is a powerful motivator to learn the L2 because of the desire to reduce the discrepancy between our actual and ideal selves’ (Dörnyei & Chan, 2013, p.438).
- *Ought-To L2 Self (OTS)*: ‘concerns the attributes that one believes one ought to possess (i.e., various duties, obligations, or responsibilities) in order to avoid possible negative outcomes’ (Dörnyei, 2009, p.29) and which may ‘bear little resemblance to one’s own hopes or desires since these represent someone else’s vision for the L2 learner in question and thus they concern an ‘imported’ image of the future that the learner will then internalize to some extent’ (2015, p.88).
- *L2 Learning Experience (L2LE)*: ‘concerns situation-specific motives related to the immediate learning environment and experience’ (Dörnyei & Chan, 2013, p.438).

2.1.2.1. L2MSS Studies

Since reviews and analyses of L2MSS studies have been numerous, a brief overview of general findings is presented here, followed by a more in-depth examination of the implications of the model and areas for development.

In 2015, Boo, Dörnyei and Ryan conducted a review of over 400 L2 motivation studies published between 2005-2014 during which there was ‘an unprecedented boom’ in research conducted in the field (p.145). The studies involved seven main theories, of which the L2MSS was the most prominent, seeing its popularity rise, in particular, from 2011 onwards. They suggest that the popularity of the model stems from its versatility, and that the validity of the model—in part or as a whole—has been supported by numerous studies. They identify emerging trends in the realms of both mixed-methods—though quantitative studies still dominate—and the study of more diverse language learning contexts—though the study of college learners of English is still predominant. More recently, Al-Hoorie (2018) conducted the first meta-analysis of L2MSS research, involving a total of 32 empirical studies. Al-Hoorie reported that the three components of the model (IS, OTS, and L2LE) were each significant predictors of criterion measures, such as Intended Effort ($r_s(\text{IS})=.61$, $\text{OTS}=.38$, and $\text{L2LE}=.41$) (p.722). This study offers a useful overview of the support for each of these components, as is detailed and expanded upon below.

2.1.2.1.1. *Support for the L2MSS*

The Ideal Self in particular, in addition to being highly correlated with integrativeness, has been identified as a reliable predictor of effort. In some cases, it explained over 40% of the variance in the criterion measures for motivation (intended effort or motivated behaviour) (Dörnyei, 2015). By way of example, Taguchi, Magid, and Papi (2009), investigating learners within three Asian countries (Japan, China, and Iran), reported that the IS predicted effort across all three contexts directly, and indirectly through learners’ attitudes to learning English. Similarly, in a study of the relationship between learners’ emotions and L2 self-concepts amongst 524 teenage learners of English in Iran, Teimouri (2017) observed the IS to be the strongest predictor of effort and the

only predictor of students' willingness to communicate. Numerous other studies across multiple contexts reported similar findings with respect to the IS being a strong predictor of outcome measures (see, for example: Csizér & Kormos, 2009; Ghanizadeh, Eishabadi, & Rostami, 2016; Henry & Cliffordson, 2017; Islam, Lamb, & Chambers, 2013; Ryan, 2009; Yang & Kim, 2011).

The Ought-to Self has been met with less support. Some studies (e.g., Papi, 2010) have reported low but positive links between the concept and criterion measures, whereas others (e.g., Kormos, Kiddle, and Csizér, 2011) reported that the relationship between the OTS and effort was not significant. Interestingly, Lee and Ahn (2013) reported negative impacts of the OTS on outcome measures. Blair and Azaz (2019) separated the OTS into two categories: 'others', more extrinsically sourced self-concepts; and 'own', those generated by learners themselves. The researchers report that the two variables exhibited dichotomous relationships with the outcome variable, intended effort, with OTS/own having a positive effect, and OTS/others having a negative one. One possible explanation for the diversity of these findings, as suggested by Taguchi et al. (2009), is that socio-cultural/educational factors may come into play more visibly when it comes to obligations and expectations. This often contradictory picture highlights the need to unpick the OTS variable and investigate its validity within specific contexts.

The L2LE presents a complex area of investigation since it has been operationalised and measured multifariously. A number of studies have adopted *Attitudes to L2 Learning* as a measure of the L2LE and have reported a positive relationship between this construct and criterion measures (e.g., Kong et al., 2018; You & Dörnyei, 2016; You, Dörnyei & Csizér, 2016;), reflecting 'a larger issue concerning the unspecified theoretical nature' of the L2LE (Dörnyei, 2019, p.23). More broadly, however, the L2LE can be seen to encompass numerous variables: the nature of the classroom, the course of study, the teacher, peers, learning resources and tasks, and learners' perceptions and interactions with all of these. It has often been seen as a catch-all category to encompass, as Dörnyei himself puts it, the vast array of 'situation-specific motives related to the immediate learning environment and experience' (2019a, p.21). However, Dörnyei has recently addressed the diverse, undertheorized nature of the L2LE, proposing the adoption of *engagement* as a new theoretical framework to capture elements of the learning

experience, stating that the L2LE ‘can be defined as the perceived quality of the learners’ engagement with various aspects of the language learning process’¹ (p.25). Though this narrows the breadth of the category somewhat by proposing a lens through which to measure the various factors in the learning environment, it may be argued that these variables are potentially innumerable, and more research is needed to explore the multidimensional nature of the L2LE.

Despite the ongoing prominence of the L2MSS in the field of motivation, and the general trend of support for the overall validity of the basic tri-partite model, this model faces challenges. Its reported shortcomings include: the exact nature of the IS and the OTS; the absence of measures of the Current Self (CS) and the related issue of self-discrepancy; limited criterion measures; a pervasive focus on English as the target language; and further issues with study design, all of which are discussed below.

2.1.2.1.2. *Issues and Areas for Development with the L2MSS*

2.1.2.1.2.1. *Ideal Self*

Dörnyei outlines the distinction between the two future self-concept² categories: ‘Ideal selves have a promotion focus, concerned with hopes, aspirations, advancements, growth, and accomplishments; whereas ought self-guides have a prevention focus, regulating the absence or presence of negative outcomes, and are concerned with safety, responsibilities, and obligations’ (2005, p.101). Whilst the desire to integrate usually falls comfortably under the IS category, according to Higgins (1987), instrumentality can be divided into two categories: *Instrumentality-promotion*, ‘concerned with hopes, aspirations, advancements, growth and accomplishments (i.e., approaching a desired end-state)’ and *instrumentality-prevention*, ‘regulating the absence or presence of

¹ The initial design of this research, formulated in 2015-16, conceptualised the L2LE as engagement. Instruments for measuring learners’ task-based engagement were designed, piloted, and implemented. However, issues of participant non-compliance resulted in the data being unsuitable for use and subsequently omitted from the study.

² The terminology used to refer to current and future self-concepts varies in the literature. It appears that the terms ‘self; self-guides; motivational self-concepts; self-concepts; and self-images’ are used interchangeably. This approach is adopted throughout this research.

negative outcomes, concerned with safety, responsibilities and obligations (i.e., avoidance of a feared end-state)' (Dörnyei, 2010, p.79). The former correlates with the IS, whilst the latter with the OTS.³ In practice, however, the dividing line between these categories may not be so clear, especially since, as Boyatzis and Akrivou (2006) point out, determining whether one's IS is genuinely a product of internal desires or that of society's expectations of normality and success is not easy.

Another oversight in the realm of the IS is its internal dynamics. Although empirical validity has been lent to the IS construct through studies employing factor analysis, the mechanisms behind its motivating power have, until recently, remained a *terra incognita*. Dörnyei proposed that the IS needs to meet a list of conditions in order to be achieved (2009, p.20). According to these, the IS should be:

- ✓ Available to the learner,
- ✓ Different from the current self,
- ✓ Elaborate and vivid,
- ✓ Plausible,
- ✓ In harmony with the learner's social environment,
- ✓ Should not be regarded as comfortably within reach,
- ✓ Regularly activated,
- ✓ Accompanied by procedural strategies, and
- ✓ Offset by a counteracting feared possible self.

Examinations of these conditions, however, have been limited. Hessel (2015) addressed this omission to some degree through her study of the motivations of 97 German learners of English. She proposed a redeveloped questionnaire which included a subset of items beneath a series of opening statements concerning a specific future L2 self. These sub-items included measures of *desirability*, *accessibility*, *plausibility*, and *effort* on an 11-point scale, with positive and negative verbalised endpoints (Figure 2.3).

³ It is also potentially the case that OTSs incorporate traditionally integrative orientations, as opposed to solely instrumental ones: for example, in situations where one is obliged to communicate and function in a foreign community, such as in cases of forced migration.

Hessel observed that accessibility (the frequency of students' imaginations of themselves as fluent English speakers) emerged as the largest predictor of effort amongst participants ($t=5.91$; $p<.001$) (p.111), and that when entered into the model simultaneously accessibility, desirability and plausibility explained 37.7% of the variance in effort (p.110). This suggests that adopting a multidimensional approach to operationalising the construct is necessary.

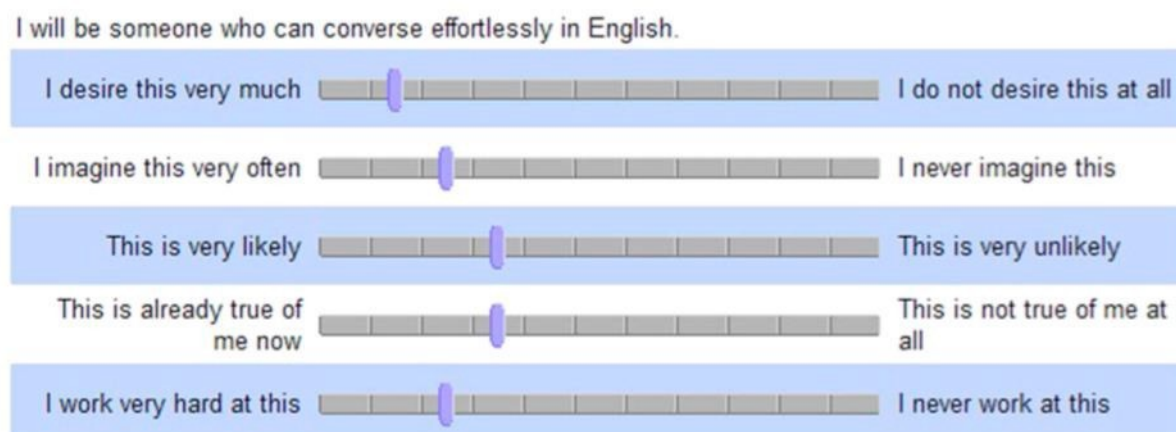


Figure 2.3. Hessel (2015)'s Motivation Survey Items

In addition, Hessel recommends the exploration of other potential IS properties, including the degree of elaboration and vividness of one's future self-concept to determine the effect of this on the strength and workings of the IS. Although she utilises only a single opening statement, Hessel suggests that different self-concepts can be measured including, for example, future-selves associated with proficiency in different skills. This would avoid another common criticism of L2MSS measures, in which only spoken skills are investigated.

2.1.2.1.2.2. Ought-to Self

Another central criticism of the L2MSS pertains to the concept of the OTS. In a number of studies it has failed to emerge as a distinct factor, for example in Lamb's (2012) study of teenage learners of English in Indonesia. Kormos et al.'s (2011) study also failed to find a significant link between the OTS and outcome variables. In others, however, it did emerge as a significant contributor to motivated behaviour

(e.g., Taguchi et al., 2009). Kormos et al. suggest: 'The comparison of our results in Chile and in Hungary with the outcomes of studies in North-Asian contexts (Taguchi, Magid, & Papi, 2009) reveals that in certain cultures where socio-educational factors put a great pressure on students' achievement and where foreign language education is highly exam-oriented (e.g., China and Japan), it is possible that the Ought-to L2 self has an effect on effort and persistence in language learning' (p.508). Lee & Ahn (2013) reported that, whilst the OTS correlated positively with Intended Effort, its correlation with achievement was actually negative. This suggests that the OTS may in fact be an amalgamation of multiple self-concepts, which need unravelling. In this sense, the issue of (re-)defining the OTS is one that also requires attention.

In addition to the dearth of significant positive findings regarding the OTS, a number of criticisms concerning its conceptualisation have been raised. Dörnyei originally defined the OTS: '[it] reflects the attributes that one believes one ought to possess to meet the expectations of others and to avoid possible negative outcomes in the process of L2 learning. It may bear little resemblance to one's own hopes or desires since these represent someone else's vision for the L2 learner in question and thus they concern an 'imported' image of the future that the learner will then internalize to some extent' (2015, p.88). This definition, however, raises a number of theoretical concerns. Firstly, it puts 'meeting expectations of others' and 'avoiding negative consequences' under the same heading. Although these may indeed overlap, empirical justification for combining them should be provided. Secondly, the source of the OTS image is also subject to question. 'Imported' implies that all OTS images are derived from external sources, which are then internalised. In light of these concerns, Dörnyei himself, along with Al-Hoorie, goes on to ask: 'Would it not be a more realistic position, instead, to state that the ought-to L2 self of LOTE [languages other than English] learners is more fragmented, and that in some situations this fragmentation might reach such a degree that it cannot be considered a unified self-dimension anymore?' (2017, p.460).

Higgins' (1987) original theory, in fact, distinguished between 'own' and 'other' sources of both the IS and the OTS. Teimouri (2017) developed questionnaire scales

to measure each of the four proposed constructs: the *ideal-own*, *ideal-other*, *ought-own*, and *ought-other*. He reported support for the distinction between own and other in the case of the OTS, but not the IS. This distinction was corroborated by Blair and Azaz (2019) who also integrated a more nuanced measure of the OTS into their study of Arabic learners in the United States. They distinguished between *ought-to L2 self/own* and *ought-to L2 self/others* in their scales, which they reported had opposing effects on Intended Effort, the latter having a positive, and the former a negative effect on the outcome variable. This offers a strong argument for the necessity of re-examining the structure of the OTS, both quantitatively in different language contexts, and qualitatively to illuminate the reasons for its presence or absence.

2.1.2.1.2.3. Current Self and Self-Discrepancy

Another major blind spot of L2MSS research is the lack of focus on the Current Self (CS). The original psychological theories upon which the L2MSS is based stipulate that motivation stems from the discrepancy between the current and future selves and the desire to reduce this (Higgins, 1987; Markus & Nurius, 1986). It is, by extension, necessary to measure learners' CS concepts to determine the gap between the two. In very few instances, however, has either the CS or self-discrepancy been operationalised and measured (Thorsen, Henry, & Cliffordson, 2020).

In addition to pushing the boundaries of research into the IS, Hessel (2015) also presented findings with regards to self-discrepancy. The fourth scale item in her instrument was designed to measure the perceived present-future self-discrepancy (Figure 2.3). In confirmation of the self-discrepancy theory, though at odds with some of Higgins' original beliefs, self-discrepancy emerged as a significant predictor of the criterion measure (intended effort), most strongly so when the perceived gap between current and future-selves was more moderate. Whilst a wider gap was understood by Higgins to promote more motivated behaviour, Hessel proposes the importance of *plausibility* in the motivating capacity of the IS: the more achievable a future self-concept is, the more effort one is likely to expend in achieving it. Although *plausibility* as a sub-scale item did not emerge as a significant predictor of effort, it appears that indirectly,

through a learner's perception of the likelihood of attaining their future self-concept by investing effort and energy in the present, it forms an integral part of the internal workings of the IS. While the inclusion of such a measure is a welcome, and indeed a long-overdue feature to studies in the field, the construct is only measured via one scale item. Furthermore, as the study is cross-sectional, no information is provided as to the dynamics of self-discrepancy. Another issue with this particular item is its confinement to 'perceived present-future self-discrepancy'. According to Higgins (1987), this construct is not one which learners are consciously aware of and therefore cannot easily report themselves. What is needed, rather, is a way of empirically calculating the gap between learners current and future selves, of which learners are conscious, and can provide.

According to Higgins (1987), there are three ways the discomfort caused by a discrepancy in selves can be addressed: the enhancement of the CS in the direction of the IS; the modification of the IS to bring it more in line with the CS; or reducing the salience of the discrepancy, for example through avoidance strategies. This might seem straightforward to capture; however, there is the added dimension of what could be termed 'optimal distance'. The size of gap between the Current-Ideal selves determines the power of the discrepancy to motivate. 'Motivational intensity will be sustained only in so far as the target—the potential self—also shifts in a forward or upwards direction... If over time the potential self doesn't keep pace with developments in the actual self, motivational energy will deplete as the discrepancy shrinks' (Thorsen et al., 2020, p.585). Dörnyei and Ryan themselves stipulate that in order to be motivating the IS 'should not be regarded as comfortably within reach', but should still be plausible (2015, p.92). How this is calculated, whether there is an empirical range for this distance, whether it is relative to the scope of an individual's self-timeline, or whether it is domain specific remain unclear. Changes in the size of the self-discrepancy could offer valuable insight into motivational changes longitudinally, particularly given the observed trends of reductions in motivation over time (Williams, Burden & Lanvers, 2002) and explain why the IS varies in its predictive power of outcome variables in different studies.

As aforementioned, there is little precedence in motivational literature for

operationalising the CS-IS discrepancy. In Higgin's (1987) own research, the researcher did not measure the size of the discrepancy; instead, he identified matches or mismatches between participants' current-future self-concepts. This approach was replicated by MacIntyre, Mackinnon and Clement (2009), who identified four categories of (mis)matches between CS and IS concepts and reported that learners who exhibited a CS and IS concept in a specific domain were more motivated than those who had only an IS image.

Of the few studies that have attempted this analysis is Thorsen et al. (2020). In their study of learners of English in two secondary school grades (grades 7 and 9) in Sweden, the researchers examined the relationship between learners' CS and IS and the outcome variable, Intended Effort, through use of Structural Equation Modelling (SEM). Self-discrepancy was measured using a nested factor model. The results of their analysis revealed, firstly, that the explanatory power of the IS was lower for students in grade 9 than grade 7. In addition, the difference between the CS and IS variables for the grade 7 cohort was significantly larger than the grade 9 one, suggesting a decrease in motivation over the school trajectory. This discrepancy variable was reported to account for approximately 15% of the variance in the criterion measure for grade 7 but did not account for any for grade 9. The researchers conclude that a smaller self-discrepancy amongst grade 9 accounts for the lack of variance explained in the Effort variable. They explain the implications of their findings: not only does the reduction in discrepancy appear to negatively affect students' motivation, but that changes in current and ideal selves are not always aligned, the former may change substantially more than the latter. 'In this light, measurement of the target construct alone becomes problematic since the element that may be most influential in determining motivational intensity, and which is most susceptible to change, is obscured' (p.594). The researchers make the pertinent point that the examination of the dynamics of learners' motivational selves both in terms of their multifaceted nature and changes over time fits within the Complex Dynamic Systems trend of research which aims to take a finer-grained approach to examining the mechanisms of motivation by examining, in detail, features of both the participant and the learning environments (Dörnyei, MacIntyre & Henry 2015) (see Section 2.1.3).

One important area of expansion when it comes to developing Thorsen et al.'s (2020)

approach is to implement a longitudinal design which incorporates measures of IS concepts. Since Thorsen et al.'s study was cross-sectional, conclusions drawn as to the reduction of motivation across time need to be corroborated. The researchers suggest that 'in school contexts where English is highly-constituent in young people's lives, self-revision trajectories may be U-shaped' (p.595). It is also possible that these trajectories may be found within grades, as well as across them, demonstrating the need for within-grade longitudinal studies. The pedagogical implications of Thorsen et al.'s findings are significant. If the discrepancy between CS-IS is decreasing, but Ideal Selves show little change, this can only be explained by increases in the CS, which would be expected over the course of learners' education. Since the decrease in the discrepancy is linked to lower levels of Effort, this suggests that to maintain motivated behaviours across grades, the IS should be simultaneously increased, in order for the discrepancy to retain its motivating power. Henry and Cliffordson (2017) suggest that teachers should dedicate effort to promoting the development and expansion of learners' IS through, for example, the vision-generating techniques proposed by Dörnyei and Kubanyiova (2014). It may be the case, although the researchers do not specify, that learners' IS are largely concentrated on the end point of their educational trajectories, rather than on long-term self-concepts. The power of such long-term goals, however, is not to be underestimated. According to Nakamura (2019), 'psychologically distant events are still highly motivational as they embody richer aspirations than psychologically proximal selves' (p.112), thus presenting rich ground for effective focus and development in the classroom.

2.1.2.1.2.4. Criterion Measures

A further criticism of L2MSS studies to date is the predominance of Effort or Intended Effort (IE) as the criterion measure. Whilst IE does constitute a subjective measure of potential future motivated behaviour, it does not capture objective outcomes of language learning. A small number of studies have examined the effects of self-concepts on measures of achievement. Kim and Kim (2011), in their study of secondary school students in Korea, reported that learners' IS did not predict academic grades, noting: 'being motivated by developing a vivid ideal L2 self through a dominant visual preference seems to be irrelevant to the level of academic achievement' (p.36). In

Indonesia, Lamb (2012) reported that the IS did not predict proficiency as measured by a C-test. Papi and Abdollahzadeh (2012), researching students in Iran, reported that the IS did not predict classroom behaviour. Unexpectedly, Moskovsky, Assulaimani, Racheva, and Harkins (2016) reported a negative relationship between the IS and language proficiency, suggesting the link between self-concepts and actual achievement is not only under-examined, but also not a straightforward one.

Al-Hoorie (2018) adds further support to this argument, concluding, in his meta-analysis of L2MSS studies: 'The results showed that the three components of the L2 motivational self system (the ideal L2 self, the ought-to L2 self, and the L2 learning experience) were significant predictors of subjective intended effort ($r_s=.61$, $.38$, and $.41$, respectively), though weaker predictors of objective measures of achievement ($r_s=.20$, $-.05$, and $.17$)' (p.722). It is therefore evident that further investigation of the interaction between self-concepts and objective language learning outcomes is necessary.

2.1.2.1.2.5. Languages Other Than English (LOTes)

A major area for expansion in the field of motivation studies is language focus. Theories can only be empirically generalisable if tested and applied to different language learning contexts. In their review, Boo et al. (2015) reported that almost three quarters (72.67%) of papers focussed on the study of L2 English. The issue of heritage learners is also one that has received little attention (Comanaru & Noels, 2009).

Mendoza and Phung (2019) conducted a critical research synthesis of 30 studies examining motivation to learn LOTes, all of which use the L2MSS. Unlike Al-Hoorie's (2018) meta-analysis, this incorporated qualitative studies, and looked largely at the aims and foci of the studies involved, rather than attempting to analyse their empirical results. The researchers identified German, French, and Spanish as the most prominent languages studied in an L2 context, followed by Mandarin. Of the 30 studies, three focussed on Arabic (Anyia, 2001; Kong, Han, Kim, et al., 2018; Thompson, 2017).

Referencing the work of Watt and Fairfield (2008) amongst others, Dörnyei and Al-Hoorie (2017) also briefly discuss the role of religion in language learning. As the former

articulate, the learning of sacred languages is often linked to religious practice, since a number of faiths integrate the original languages of scriptures into their rituals and practice, including Islam (Classical Arabic), Judaism (Hebrew), and Hinduism (Sanskrit). Although English has been examined in the context of Christian missionaries (Lepp-Kaethler & Dörnyei, 2013), studies examining the motivations behind the unique combination of LOTEs for religious purposes are scarce. In the context of the L2MSS, religion may pose a particular issue in terms of the overlap between the IS and OTS. Are religious obligations considered part of one's ideal or ought-to self-concept, and if one feels compelled by a divine or sacred power, can this source be said to be internal or external?

Another pertinent issue in the realm of LOTEs is that of diglossia, and its interaction with motivation. Described as 'a particularly fertile environment for the emergence of powerful unconscious attitudes' (Dörnyei & Al-Hoorie, 2017, p.464), diglossia refers to the co-existence of higher and lower varieties of a language used in different, though often overlapping contexts in a community (Ferguson, 1959). This ties in with Nakamura's (2019) conclusion from her qualitative study of Japanese language learners, in which she describes the necessity of 'examining various domains of learners' ideal L2 selves, as well as speech genres used in the domains' (p.112). The assumption that each language is a monolith which can be used across all contexts is a misleading one, particularly in cases where diglossia, or multiglossia, is the norm.

2.1.2.1.2.6. Study design issues

Of the 335 empirical papers reviewed by Boo et al. (2015), 73 (21.79%) employed mixed-methods design. Though this is significantly less than the 178 quantitative studies in the review, it does, according to the researchers, represent a 'paradigm shift' in the field, which incorporates more nuanced analyses of motivation as a dynamic and complex construct (p.153). At the same time, they caution that, despite the promising future potential of mixed-methods research, the majority of studies currently purporting to employ this approach fall short of maximising the full scope of mixed-methods design, involving instead a 'rather superficial mixing of relatively independent qualitative and quantitative components' (p.156). This highlights the need for the full integration of

quantitative and qualitative methods in order to harness the power of such study designs.

In terms of participants, according to the same review, the UK context is also under-represented in these studies, with Japan, the United States, and China constituting 35.52% of the countries in focus. The largest participant group is university students, presumably given the ease of access and the comparatively simpler ethics requirements needed for adult learners. Research on secondary school students, 'which might be considered the most suitable age group for instructed SLA' (p.151) is considerably lacking. The few studies which have focussed on secondary school learners, such as Graham (2006) and Taylor and Marsden (2014), reported low levels of motivation in this age range. Similarly, Thorsen et al. (2020), in their cross-sectional study of learners in Sweden, observed that levels of motivation drop in the middle of secondary school only to potentially rise again in later grades, explained by fluctuations in self-discrepancy. This indicates the need not only for a focus on the UK learning context, but also on the potential fluctuations of secondary school students' motivation through longitudinal research designs.

2.1.3. Complex Dynamic Systems Theory

Despite the prominence of the L2MSS, a new paradigm shift has taken root in the field: Complex Dynamic Systems Theory (CDST). While prior to this phase studies generally offered snapshots of learner profiles and teased out linear associations between supposedly static variables to construct an image of the factors at play, CDST attempts the opposite: a deconstruction of the variables at the granular level to examine the dynamics that exist within and between them. As the name might suggest, CDST is not an easy framework to define. In fact, a noteworthy compendium of work on CDST theory and research, edited by Dörnyei, MacIntyre and Henry (2015), offers no definition for the theory in its introduction, nor in its first chapter, making it, seemingly, as elusive to capture as what it is trying to capture.

According to Hiver and Larsen-Freeman (2020), 'CDST is a poststructural ecological theory' (p.286) through which researchers understand 'the relations that connect the

components of a complex system' (p.287). It is not a new model of motivation, but the adoption of an ontological approach—a way to understand the variables within a system and their inter-relations, not exclusive to a particular field. Dörnyei and Ryan (2015) offer this insight into the workings of CDST:

When viewed from a complex dynamic systems perspective, not only are individuals a product of the constant interactions between their various individual attributes and contexts, but those attributes themselves are also multi-componential in nature and make up a holistic dynamic framework. (p.11)

At this micro-level of focus, multiple components are thus at play, reciprocally and simultaneously: the learners and the mechanisms and attributes at play within them individually; the context; the interplay between individual and context, and also between learners; and the time factor, since all previous component categories are subject to longitudinal fluctuations. A crucial addition in the context of Arabic is the language focus of motivational efforts itself. Rarely is any attention given to the dynamics of the language being studied, or how it is conceptualised by its learners. Arabic presents a unique case in this respect, given the multiglossic nature of the language and its generally over-simplified portrayal in the classroom and in research.

Whilst there is undeniably a need for this in-depth, fine-grained approach to the dynamics internal to the learner and at play within the environment, as is clearly demonstrated by the diverse findings of the L2MSS when applied in different learning contexts, there have been many challenges to the application of CDST in real-life research cases. Three of these are highlighted by Dörnyei et al. (2015). The first is the question of how to make predictions if linear relationships are no longer the focus. The second concerns the extent of granulation and interconnection—where does the system end? If everything is interconnected, where should examination of a variable stop? Is defining a 'system' even possible? And thirdly, the challenge of identifying research approaches and instruments which capture the dynamics at play, rather than linear relationships alone. There is clearly no one 'right' way to go about conducting research under the umbrella of CDST. What is important, however, is that the finer-grained

approach that it offers provides a way to address the inconsistencies, contradictions and oversights that L2MSS studies, and research into Arabic motivation in general, have proffered thus far. Further, it avoids the danger of over-generalising results to other diverse and unique systems.

2.1.4. Summary

Although the introduction of the L2MSS represented a step forward in motivational research by marrying theories from educational and psychological literature, significant areas within the framework require development from both the perspective of the originating theories, and more recent expansions into CDST. Support for the validity of the OTS construct is fragile, with some studies finding only tentative evidence in its defence. Further investigation is thus required in this domain. Measurements of the CS, in order to examine its relationship to future self-concepts and analyse the discrepancies between the two, are also lacking. In addition, there is also a need to address the methodological limitations of previous studies. Longitudinal measures of motivations are needed to capture fluctuations in what is considered a dynamic construct, subject to influence from environmental factors. The effective incorporation of qualitative design elements into the measurement of motivation, such as open-ended items and interviews to allow participants to explain and expand upon their responses, is also vital to facilitating in-depth analysis of learners' self-concepts and their internal mechanisms. Finally, there is a clear need for the application of the theory to language contexts other than English in order to assess the validity of the model and the generalisability of previous research findings, particularly in LOTE-learning contexts. Arabic presents a context ripe for such fine-grained analyses, particularly given the unique multi-dimensional nature of the language itself, its ties to religion which may pose a potential challenge to the conceptualisation of the IS and OTS categories, and the growing uptake in the study of the language internationally.

2.2. Arabic

As the official language of 23 countries stretching from the borders of Iran to the western coast of North Africa, Arabic is spoken by approximately 300 million people (Kendall, 2014). It is also a liturgical language, used by over a billion Muslims worldwide for daily worship and religious rituals. The Quran, the sacred text of Islam, is preserved in its original language. As such, many Muslims consider it a duty to learn Arabic in order to understand this primary religious source and critically engage with their faith (Scott-Bauman & Contractor, 2012).

In addition to this religious impetus, a vast array of political, economic, security, and social factors, including geopolitical changes in the Middle East have increased the prominence of Arabic as a world language, fuelling calls for it to be taught across Europe and America (Al-Batal, 2007). Other factors that have played a role in the spread of Arabic include the effects of globalisation and mass movement of peoples internationally. Arabic has been identified as a 'Strategically Important and Vulnerable Subject' in the UK higher education context (HEFCE, 2011), and was identified by the British Council as second in a list of the top ten languages the UK needs most over the next 20 years (Tinsley & Board, 2013). In Britain, Arabic is the fifth most widely used 'community language' (Bernasek & Canning, 2009). According to the UK census, over 240,000 people identified as Arab. Five percent of the population identified as Muslim, 74% of whom were from non-Arab backgrounds (Office for National Statistics, 2011).

2.2.1. The Arabic Language

The Arabic alphabet is composed of 28 letters (illustrated in Figure 2.4), which are written from right to left. Each letter has a consonant sound, some of which have no corresponding English equivalents. These letters are joined together in writing (as is demonstrated in the text below the alphabet, Figure 2.4), with the vast majority presenting in modified forms depending on their position in the word. Vowels are denoted by the use of diacritics, known as *ḥarakāt*, written above or below the letters. These are used to denote the internal vowelising of words, as well as grammatical case endings.

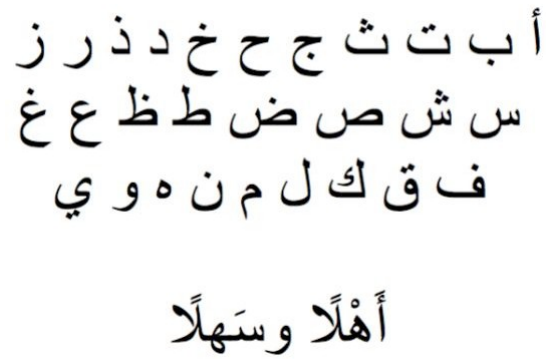


Figure 2.4: The Arabic Alphabet and Script

In the majority of written Arabic texts, these diacritics are almost entirely absent. During their schooling, Arab speakers learn to read without the aid of diacritics, relying on their lexical knowledge, context, and familiarity with morphological patterns to pronounce and understand texts correctly (Abadzi & Martelli, 2014). An exception to this norm is the Quran, which is fully vowelled in order to avoid mispronunciation or misinterpretation.⁴

The morphology and syntax of the language follow a highly regular system of verb conjugation, nominal derivation, and syntactical structure. Arabic is based on a root system, whereby thousands of trilateral (three letter) roots can be manipulated in conjunction with morpho-syntactic rules to access and produce a large volume of vocabulary. The root *d-r-s* درس (to study), for example, can be conjugated to form the imperfect tense (Table 2.1):

⁴ The Arabic script at the time of the revelation of the Quran (610-632 AD) was, in fact, written without dots or diacritics. The introduction of these symbols took place between the later seventh and eighth centuries, largely due to the need to facilitate access to religious texts by the growing non-Arab Muslim communities (Versteegh, 2014).

Table 2.1. Conjugation of ‘to study’ in the imperfect tense			
English	Arabic	English	Arabic
He studies	يَدْرُسُ	You (dual masc.) study	تَدْرُسَانِ
They (dual masc.) study	يَدْرُسَانِ	You (pl. masc.) study	تَدْرُسُونَ
They (pl. masc.) study	يَدْرُسُونَ	You (sing. fem.) study	تَدْرُسِينَ
She studies	تَدْرُسُ	You (dual fem.) study	تَدْرُسَانِ
They (dual fem.) study	تَدْرُسَانِ	You (pl. fem.) study	تَدْرُسْنَ
They (pl. fem.) study	تَدْرُسْنَ	I study	أَدْرُسُ
You (sing. masc.) study	تَدْرُسُ	We study	نَدْرُسُ

2.2.1.1.1. Perceptions of Arabic

Arabic is often considered to be a very difficult language to learn as an L2 (Belnap, 2006), and has been categorised as Level IV by the US Foreign Service Institute (FSI), along with Chinese, Japanese, and Korean—all described as ‘exceptionally difficult for native English speakers’ (US Department of State, n.d.).

A plethora of reasons are cited in support of this perceived difficulty of Arabic. Orthographically, the need to learn as many as four possible shapes for each letter, coupled with the lack of transparency in reading the script given the absence of diacritics (omitted in the majority of texts), poses a challenge in terms of phonetic decoding for early readers of modern texts. The absence of diacritics also increases the number of heterophonic-homographs in the written lexicon, further complicating the reading process (as illustrated in Table 2.2).

Table 2.2. Heterophonic-homographs from root عَقَدَ			
Meaning	Transliteration	Without diacritics	With diacritics
He knotted	‘aqada	عقد	عَقَدَ
It was knotted/ it was convened	‘uqida	عقد	عُقِدَ
It thickened	‘aqqada	عقد	عَقَّدَ
It was thickened/ made more complex	‘uqqida	عقد	عُقِّدَ
A decade	‘aqd	عقد	عَقْدَ
A contract	‘aqd	عقد	عَقْدَ
A necklace	‘iqd	عقد	عَقْدَ
Knots	‘uqad	عقد	عُقْدَ

In fact, studies suggest that cognitive processing time in Arabic reading, even amongst native speakers, is slower than in other languages, such as English; a result of this orthographic complexity (Eviatar & Ibrahim, 2014). In theory, according to Abadzi and Martelli (2014), learners should be able to acquire the ability to phonetically decode vowelised Arabic texts in approximately 100 days. However, as these researchers note:

[The] removal of the vowels makes the meaning opaque and creates dependency on neighboring words and context.⁵ Unlike other languages, even adult experienced readers must rely on context. Readers must retain alternatives in working memory to make sense of text. The alternatives compete with the meaning of the passage. As texts become more complex, the working memory load gets an extra burden. Simultaneously the readers must decipher the meaning of standard Arabic expressions that may be unclear. All these decisions may be reducing the amount of time that students have to decide about the meaning of a text. (p.3)

A further difficulty for learners of Arabic (particularly native English speakers), is the fact that, as a Semitic language, Arabic shares very little with English with respect to vocabulary when compared to languages such as French and Spanish (England, 2014). In addition to the lack of cognates, the size of the lexicon and the number of near synonyms—words which have similar but not interchangeable meanings—present a challenge for learners, who are faced with the task of accumulating a vast array of words often to express even simple concepts (Anderson & Suleiman, 2009). Other hurdles include the challenge of pronunciation, and elements of Arabic morphosyntax.

However, while there may appear to be general agreement on the difficulty of the Arabic language for English speaking learners, within the research community there is a degree

⁵ There is some debate in the field of Arabic pedagogy as to the utility of removing vowels. According to a study by Hansen (2010), the suppliance of internal short vowels did not aid beginner-intermediate learners in either reading speed nor comprehension. This is in contrast to earlier studies, such as Abu Rabia (1999), which demonstrate the opposite. Hansen's explanation for these findings is that the addition of vowels adds to the already heavy cognitive burden early learners face in reading Arabic. Alhawary (2017), however, points out that information is not supplied by Hansen as to the nature of these learners first exposure to vocabulary items in question. If they had first been learned without vowels, then the addition of diacritical markings may well be overwhelming. Studies are therefore required into the effects of adding/ removing vowels from words for students who have learned the words with and without these diacritics, so as to compare the effects.

of push-back against this claim, which, some argue, can often translate into a self-fulfilling prophecy. Stevens (2006), for example, takes great issue with the ranking of Arabic as a Level IV language. He compares Arabic to other languages on the FSI on the basis of language-relatedness, grammar (inflectional and derivational), vocabulary and lexicon, and the writing system. In each category he identifies languages that outweigh Arabic in terms of the difficulty of their linguistic features. For example, whilst Arabic's distance from English certainly contributes to making it more challenging for English speakers to learn than Romance or Germanic languages, when compared to Indo-European languages, such as Farsi and Urdu which are classed as Group II, according to Stevens, these languages appear to have just as little overlap with English in terms of their lexicon. Although he clarifies the fact that learning Arabic is certainly not to be dismissed as 'easy', Stevens concludes that 'when seen in context of other languages it is not as difficult, relatively speaking, as is commonly believed' (p.62).

This sentiment is echoed by other researchers, such as Belnap (2014) who neatly summarizes the issue:

Arabic has a reputation among Americans for being a difficult language to learn. The novice is wowed by the script, but it is easily learned. Grammatically, it pales in comparison with some of the world's other languages. In fact, nothing about it stands out as being clearly more difficult for the learner than comparable aspects of many other languages that do not enjoy, or perhaps better, are not cursed with the same notoriety. It is probably no coincidence that Arabic is widely held by Arabs to be the most difficult language in the world. A good deal of experience suggests this belief about the difficulty of learning Arabic is regularly translated into a self-fulfilling prophecy. (p.175)

However, there are undeniably several issues which do present challenges to learners of Arabic as a foreign language. One particular challenge—and perhaps the largest pedagogically—is the socio-linguistic features of the language, namely the existence of multiple varieties of Arabic. The reality of authentic Arabic usage is nuanced and complex and presents considerable obstacles to learners and teachers alike.

2.2.1.2. Diglossia

The term 'Arabic' is arguably a misleading one, a catch-all phrase covering a number of varieties. The existence of different language varieties, usually split into 'higher' and 'lower' branches, was originally termed 'diglossia' by Ferguson (1959). The higher, usually the standardised variety, represents 'the language of power and control', in contrast to the lower, the colloquial strand, symbolically that of 'intimacy and domesticity' (Holes, 2004, p.6). In the case of Arabic, the 'higher' variety is represented by *fuṣḥā*, formal varieties including Modern Standard Arabic (MSA), and the 'lower' by colloquial varieties, *‘āmmiyyah*.

Typically, Arabic in educational settings refers to MSA, which is usually taught in isolation from any other varieties. According to studies in neuroscience, however, MSA is represented in the brain of Arabic native speakers (NS) as a second language, not as a L1 (Ibrahim, 2009). Further research by Khamis-Dakwar and Froud (2007) confirmed that standard and colloquial varieties have distinct neural lexical stores, with the former stored as a second language and the latter as participants' first.

The theoretical bifurcation between standard and colloquial varieties is now generally regarded as an oversimplification. Though the two are neuro-linguistically distinct, on a pragmatic level the interactions between the two are complex. As explained by Ferrari (2018), MSA and colloquial varieties 'are well-defined at the abstract level in the speakers' minds and the two language varieties are separated at the neuro-cognitive level, but at the practical linguistic level they are performed as one language, which is part of one Arab(ic) cultural system' (p.45). Attempts to anatomise the language have subsequently depicted it as a spectrum: 'The behaviour of most Arabic speakers, educated or not, is rather one of constant style shifting along a cline at opposite ends of which are 'pure' Modern Standard Arabic [*fuṣḥā*] and the 'pure' regional dialect [*‘āmmiyyah*]' (Holes, 2004, p.49).

In addition to the formal Modern Standard and spoken Arabic dialects is the existence of the Classical Arabic variety, the language of liturgy. The uses of these three branches, whilst in practice intercepting in different contexts, can be relatively discrete, as

illustrated in Table 2.3. This has contributed to the description of the language as ‘multiglossic’ (Badawi, 1973).

Table 2.3. Arabic Varieties (adapted from Wahba, 2006)			
Variety	Modern Standard Arabic	Classical Arabic	Dialect/ Colloquial
Usage	Media; academic texts	Reading sacred/ liturgical texts; sermons; classical literature	Day-to-day interactions
Population	Educated Arabic speakers	Muslim population (Arab and non-Arab)	Native Arabic speakers
Taught/ learned	In school/ university	In religious institutions	Acquired as L1 through natural interactions
Skills	Predominantly written/ read (spoken in formal contexts, e.g., news broadcasts/ interviews/ academic lectures)	Written/ read in religious contexts	Spoken

2.2.1.3. Standard Arabic

Fuṣḥā, or the ‘higher’ variety, covers both Modern Standard (MSA) and Classical Arabic. Classical Arabic exists in much the same form now as it did in the seventh century when the religion of Islam emerged and has ‘been protected from evolving apace owing to its sanctification as the language of the holy Quran’ (Kendall, 2014, p.3). This variety is written and may be read aloud in sermons and prayers, but it is not spoken.

MSA ‘has not been spoken as a first language (L1) for centuries or even millennia, but [...] is nevertheless the official language of all the Arab countries and is taught in schools from the first day of the first grade’ (Resto, 2013, p.433). This variety is the contemporary offspring of Classical Arabic, ‘unchanged in the essentials of its syntax but very much changed, and still changing, in its vocabulary and phraseology’ (Holes, 2004, p.5). It is seen as ‘the international language of communication among Arabs of different colloquial dialect regions’ (Allen, 2004, p.276). ‘This unified, codified pan-Arab variety of

Arabic is used for virtually all writing in the Arab world and nowadays, in its spoken form, also dominates the airwaves and the TV channels of every Arab country. As the normal medium for formal discourse, it is used for all news broadcasts, political speeches, official announcements, and—most crucially—education in every Arab country.’ (Holes, 2004, p.5), Proficiency in MSA and Classical Arabic both form part of native speaker (NS) competence in Arabic-speaking communities used in tandem with an individual’s local dialect (Wahba, 2006).

2.2.1.4. Colloquial

Colloquial Arabic, or *‘āmmiyyah*, constitutes the medium for everyday conversation. A person’s local dialect is acquired as their L1, though it is not explicitly taught in schools. This is likely due to its perception by Arabs as a ‘grammarless corruption of ‘real’ Arabic’ (Holes, 2004, p.5). Dialects, though regionally distinct, appear to have porous borders; ‘the vast geography covered by the Arabic-speaking world engenders significant divergent forms of spoken Arabic (Gulf, Egyptian, Levantine, North African and so on, as well as their many sub-groups), despite the political desires of pan-Arabism enthusiasts to play down these differences’ (Kendall, 2014, p.5) (see Figure 2.5 for an example of regional phraseological variation). Whilst NSs within these geographic regions may understand their ‘neighbours’, the variation in vocabulary, grammatical features, and pronunciation, on top of the varying local, national, and regional accents, can make comprehensibility of spoken dialects from different regions difficult. This often-borderline unintelligibility can supposedly be avoided, as Resto states: ‘linguistically conscious speakers often can make their way and understand each other in spite of the differences’ (2013, p.433) by dipping into the ‘communally owned linguistic reservoir’ of MSA (Holes, 2004, p.6).

Colloquial Arabic writing was traditionally confined largely to cartoons, informal letters, and occasionally dialogues in modern novels. As such, it generally conveyed ‘associations of the ‘domestic’, ‘homely’, ‘amusing’, and ‘non-serious’ (Holes, 2004, p.50). However, with the rise of social media, the non-standardised writing of Arabic dialects has proliferated, making it evermore necessary in the digital age for learners of the language to master it.

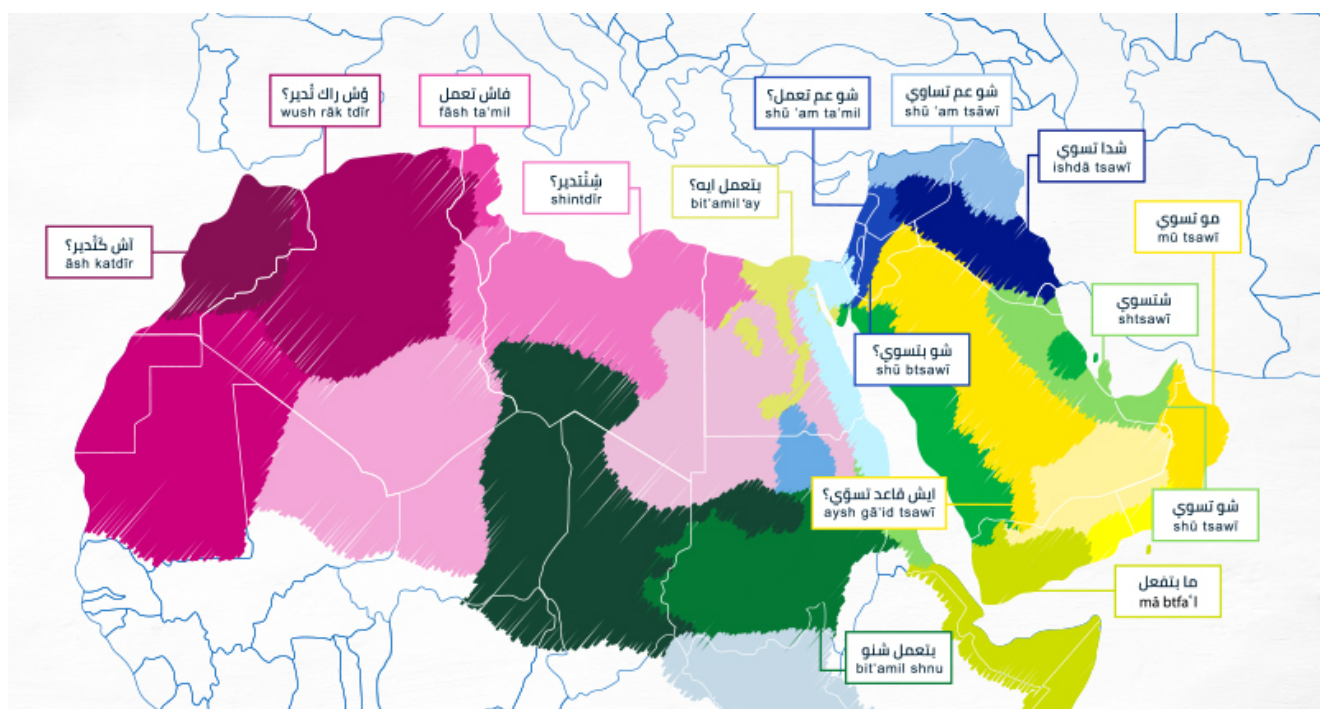


Figure 2.5. Infographic on Variance of Dialects in the Middle East (QFI, 2021)

2.2.1.5. A middle ground?

In reality, these two branches—standard and colloquial—are intertwined in everyday life for NSs, in a state of ‘constant flux’ with ‘ever-increasing leakage between the two varieties’ (Omar, 2011, p.3). A great deal of communication is conducted in a blend of the colloquial and formal varieties, often referred to as ‘elevated Arabic’, or *mutawassit*, meaning ‘middle’ or ‘intermediary’. This is seen as an ‘approach, a state of mind’ which proficient speakers should learn to master in order to replicate native-speaker practices (Scott-Baumann & Contractor, 2012, p.20) rather than as a separate branch of the language. As such, the Arabic spoken in authentic contexts consists of ‘a whole continuum of levels of possible variations which depend on many non-linguistic factors such as setting (formal or informal), topic (serious or light), linguistic skill and mastering of Classical Arabic, emotional state of speakers, number of participants in the discussion, function of the discourse, and personal relationship with the audience’ (Abboud-Haggag, 2011, p.2). In other words, Arabic native speakers ‘develop a range of diglossic competencies based on

their understanding of the interplay of SA [MSA] and QA [colloquial Arabic] in different social domains. These multiple competencies enable speakers of Arabic to perform various language activities and fulfil different roles in response to situational and functional factors' (Albirini, 2016, p.300).

2.2.2. What should be taught?

Debate continues in the Arabic practitioner and research community over the most appropriate approach to teaching Arabic, in the fullest sense of the term, to equip students with the tools to function in authentic contexts. Most institutions opt for a focus on MSA. Donitsa-Schmidt, Inbar and Shohamy (2004) offer a succinct summary of the reasons for this stance in the context of teaching Arabic in Israel:

First, experts argue that teaching spoken Arabic first may lead to inaccurate language use (e.g., Brosh, 1988), as well as to the adoption of harmful learning tactics by reverting to the use of transcribed Hebrew orthographic symbols (because the language is an oral one and cannot be transcribed). It is easier, they believe, to proceed from MSA to the spoken language, than the reverse. Second, teaching MSA provides access to the whole Arab world and not just to one country. Third, MSA is considered to be more prestigious than the spoken dialects. Finally, a decision to teach spoken Arabic also leads to major difficulties regarding the issue of choice of dialect. (p.220)

Many of the points these researchers raise, however, have been challenged. Though calls for a sole focus on *fuṣḥā* persist in many circles, consensus is growing concerning the need to integrate other varieties of the language within the classroom: '[it is] generally agreed that academics could not redefine the language they wanted to teach; they had to engage with Arabic as it was spoken and used by Arabs' (Anderson & Suleiman, 2009, p.131). Al-Batal (2018), co-author of one of the most prominently-used textbooks at university level, commonly known as *Al-Kitaab*, has recently edited a compendium of research in support of integrating dialect into Arabic language curricula. One of the aims of this work was to shift the debate away from the realms of ideology—given pervasive beliefs amongst many Arabic speakers that *fuṣḥā* is the 'correct' or 'superior' variety and

the only one which should be taught—to the realms of pedagogy and discussions on what would most faithfully cater to students’ learning needs. It offers a strong critique of the MSA-only model and calls for a broadening of the scope of Arabic teaching. A number of different approaches to diversifying the teaching of Arabic in classrooms have emerged over the past decades. Where dialects were originally incorporated pedagogically this tended to be in succession to MSA, usually via means of a year abroad. Less frequently, learners have been taught a dialect first, followed by MSA, a process intended to replicate the experience of L1 Arabic learners (Qafisheh, 1972). Growing interest has focussed on an integrated approach, of which the *Al-Kitaab* textbook is a prime example, in which the two varieties are taught simultaneously (Younes, 2006). Most recently, however, Trentman and Shiri (2020) have proposed a variationist approach which advocates a multidialectal pedagogical model in which learners are exposed to multiple dialects alongside MSA, with emphasis on developing metalinguistic awareness of the similarities and differences between these varieties, enabling learners to navigate through real-life situations in which these varieties are combined.

2.2.2.1. Pedagogical Mismatch

The pedagogical challenge presented by this complex linguistic situation is a significant one (Nielsen, 2015). In most educational institutions in non-Arab countries, the standard variety is given preference, often taught in isolation from any mention of the colloquial (Donitsa-Schmidt et al., 2004). Whilst learning MSA is certainly vital for students to be able to function in Arabic-speaking contexts, ignoring the sociolinguistic reality of Arabic use comes at a cost: ‘a top Arabic graduate thoroughly versed in core Islamic texts would likely still struggle with the transcript of an everyday conversation, let alone be able to join in seamlessly, unless s/he already had significant exposure to the region in question’ (Kendall, 2014, p.6). The labelling of MSA courses as simply ‘Arabic’ raises the issue of both the authenticity of the material being taught, and of transparency. Arabic courses ‘remain opaque over the type and level of the Arabic they teach, [and] this results in a mismatch between learning outcomes and employer (and indeed student) expectations’ (p.6), which could adversely affect learners’ motivations.

2.2.3. Teaching Arabic in the UK

The study of Arabic in Britain can be traced back to the twelfth century, during which Arab Islamic thought is said to have experienced the peak of its scientific and philosophical achievements. Arabic has been taught formally in British universities since at least 1632 when the Sir Thomas Adams Chair of Arabic was established at Cambridge University (Dickins & Watson, 2006). Despite this long scholarly history, provision of Arabic in schools has been, for the majority of the past four centuries, absent and although it has witnessed an emergence in recent decades, teaching is still sparse.

Approximately 4% of UK schools teach Arabic (Tinsley, 2015). The existence of the GCSE (General Certificate of Secondary Education) exam in Arabic has acted as an incentive for many schools to offer the subject, laying out the required course materials and assessment objectives to be covered. The majority of these are Muslim faith schools, in which the National Curriculum is taught within a religious ethos, along with the teaching of subjects related to the Muslim faith. A small but steadily increasing number of non-Muslim schools are offering the subject at GCSE; however, official exact figures are difficult to identify. In May 2018, the British Academy launched a pilot project aimed at identifying all institutions where Arabic was taught. The Arabic Mapping Project provided an interactive map (Figure 2.6) and downloadable spreadsheet with data on the nature of each institution. This spreadsheet lists 147 secondary schools and 31 'all-through' (combined primary and secondary) schools offering the language. Of these, 99 are identified as Muslim faith schools. In the notes section of the spreadsheet, it is stated that a number of the total 178 schools listed do not offer full Arabic GCSE teaching provision, rather they either enter candidates or offer it as an extra-curricular subject. The mapping project presents a useful resource for Arabic language researchers; nevertheless, the method for updating the database is not detailed, and the results suggest an incomplete picture of the landscape of Arabic teaching provision.

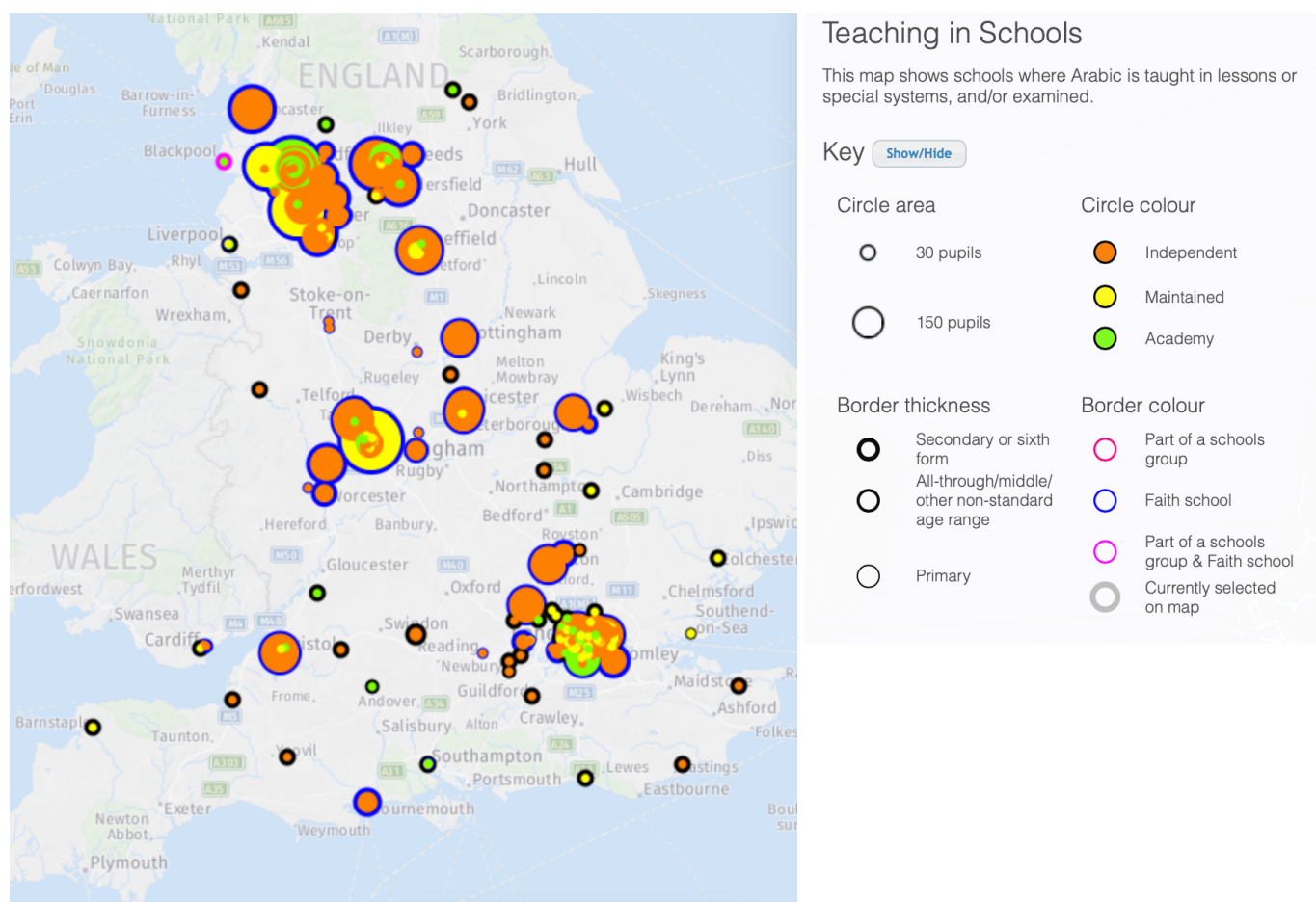


Figure 2.6. Schools offering Arabic in England and Wales (British Academy, 2018)

2.2.3.1. The Arabic GCSE

In their overview of the state of, and obstacles to, teaching Arabic in the UK, Gough and Calderbank (2019) identified the GCSE Arabic exam itself as a significant challenge. In 2015, proposals were put forth to scrap a range of community languages, including Arabic, at both GCSE and A-Level. After intense campaigning by teaching professionals, academics, and other stakeholders, the Education Secretary announced the government's U-turn on the subject. This ushered in developments in the GCSE and A-Level specifications to ensure that the language courses 'meet the rigorous standards put in place by government' (Department for Education, 2016).

The past two decades have seen Arabic GCSE entry figures grow threefold, from 1,000 in 1995 to over 4,000 in 2017 (Tinsley & Board, 2017). This increase stands in stark contrast to the plummeting uptake of Modern Foreign Languages (MFLs), which fell by

34% over the same period. Although Arabic constitutes only 1% of the overall figure for GCSE subject entries, in 2013 it became the eighth most popular language in the UK (Tinsley, 2015). However, more recent figures suggest a sharp drop in uptake, with a 28% fall in numbers between 2019-2020 (Tinsley, 2020). This may be explained by the havoc wreaked upon the school system by the coronavirus pandemic, but it may also be due to the changes in the course content and grading system, which came into full effect in 2019, and arguably made the subject inaccessible to many students.

According to Ofqual's website, in 2020 2,895 students took the Arabic GCSE. The distribution of grades awarded is found in Table 2.4.

Table 2.4. Grade Distributions for GCSE Arabic—Summer 2020 (adapted from Ofqual, 2020)										
	9	8	7	6	5	4	3	2	1	U
Number	670	545	465	320	390	245	150	50	20	40
%	23.20	18.80	16.10	11	13.50	8.40	5.20	1.80	0.70	1.30

What remains unclear from these official figures is the number of students sitting exams as independent candidates (often native/heritage learners), who have studied in supplementary schools or independently, and those who have been taught the subject as part of their school curriculum. The published figures also do not offer an indication of how many students took the full or short course (see Section 2.2.3.1), nor which tier they were entered for.

Teachers of Arabic have raised the concern that the exam caters to native and heritage speakers and that the level of language required to achieve a top grade reflects this (Gough & Calderbank, 2019). A comparison of grades for Arabic with the grade charts for other community languages (e.g., Chinese; Polish; Italian) shows a similar pattern of distribution: a preponderance of top grades amongst candidates, with the bar graphs exhibiting patterns of extreme positive skewness (Ofqual, 2020). This stands in contrast to the bell-curve distributions of other mainstream subjects, adding support to the argument that these patterns may be attributed to high levels of native/heritage learner candidates taking the exams, who achieve higher grades.

2.2.3.1.1. *Specification*

The GCSE is usually taken in Year 11, at the end of Key Stage 4. Examinations and course materials are offered by the Pearson exam board (formerly known as Edexcel). The course typically takes two years to complete, with students commencing GCSE-focussed study in Year 10.

The exam board offers both a short course and full Arabic GCSE course. The full course covers four units: Listening and Understanding; Speaking; Reading and Understanding; and Writing. For the short course, students take two of these four papers, focussing either on the spoken or written language. The Arabic GCSE was recently updated for first teaching in 2017 and first examination in 2019. It now includes a Higher and Foundation Tier, though one level must be taken across all papers. Papers are graded according to the new 9-1 system which replaced the alphabetic system of grades. A list of core vocabulary and details of the topics covered by the course are provided in the specification (Pearson, 2017). These span five themes:

1. Identity and culture
2. Local area, holiday, travel
3. School
4. Future aspirations, study and work
5. International and global dimension.

The type of Arabic that should be taught in the course is not specified explicitly in the exam board-issued materials. However, the core vocabulary and exam materials clearly suggest the Modern Standard variety is adopted. Brief mention is made of the 'standard spoken' variety in the context of the speaking and listening papers; however, a definition and the parameters of this variety are not elaborated. As was made clear by Ramezanzadeh (2016):

No guidance is provided in the mark schemes as to how to deal with instances of dialect use, either in the spoken or written units of the exam. In light of the descriptions of the varieties of Arabic and their uses hitherto

extrapolated, even a cursory glance at the topics covered in the course, all of which represent topics that would fall comfortably under the umbrella of daily interactions and thus dialect use, should set off alarm bells, raising issues regarding both transparency in the marketing of the GCSE Arabic syllabus, as well as the authenticity of the language and skills being taught.

In 2020, the researcher was commissioned to submit a report as part of the Creative Multilingualism project on an analysis of the GCSE and A-Level Arabic papers issued by the Pearson exam board, specifically the Reading and Understanding, and Listening and Understanding papers (Ramezanzadeh, in preparation). Both the legacy (pre-reform) papers (2018) and the reformed papers (2019) were examined to determine their suitability for learners of all backgrounds. Despite the statement in the GCSE specification that ‘texts for individual questions within the assessment use high-frequency language’ (2017, p.36), according to the researcher’s analysis of the verb vocabulary contained in the specification, over 50% are above the 2,000 level recommended for GCSE level of proficiency both in the legacy and reformed versions. Approximately 27% of the words in the reformed specification are above the 5,000 word band (see Appendix A.1 for a full breakdown of word frequency bands). An examination of each of the tasks in the pre- and reformed papers ‘indicates that almost all questions include multiple low-frequency words and grammatical structures which are noticeably complex’ (Ramezanzadeh, in preparation, p.10). The issue of lexical difficulty is compounded by the following statement in one of the examiners’ reports: ‘candidates must be aware of non-vocabulary words’ (Pearson, 2019, p.15). Other issues raised concern the large volume of tasks to be completed in the allotted time, the unsystematic use of diacritics, as well as the pace of speech in the listening papers, all of which contribute to the potential inaccessibility of the paper for learners of all backgrounds. Processing speeds for Arabic texts have been observed to be slower than for English, given the heavy cognitive load that decoding unvowelled Arabic scripts entails. According to Abadzi and Martelli (2014):

[T]he interaction of the perceptual and linguistic complexities may [...] turn Arabic reading into a complex multistage exercise. A reader of the Arabic script must (a) decipher the letters, (b) predict the short vowels and keep

multiple alternative words in working memory to test and decide on meaning, and (c) make linguistic sense. This process requires faster identification of words than in other scripts in order to make sense of the text, but in fact they identify them more slowly and make more errors. (p.4)

Standardizing the specification and exams for Arabic to bring it in line with other languages is thus highly problematic. Empirical evidence suggesting the inaccessibility of these exams for some learners has also been found. In a study exploring the performance of GCSE Arabic students on a number of exam tasks,⁶ Ramezanzadeh (2015) reported that heritage learners significantly outperformed non-heritage learners in both reading and listening.

Hence, whilst in theory the GCSE presents an opportunity for a diverse array of learners to learn the Arabic language, in practice the picture is more limited. At present, the majority of schools offering the language as a taught subject are Muslim faith schools, and it is not clear how well students are doing in these classes.

2.2.4. Heritage Learners of Arabic

Originating in Canada in the 1970s (Cummins, 2005), the term 'heritage language' and the concept of heritage learners (HL) has become synonymous with 'community' language (Murphy, 2014). There is still much debate over the exact definition of a HL. Bale (2010) proposes a framework for categorising different definitions by divides them into ancestral- and proficiency-based approaches.

The proficiency-based approach identifies as HLs individuals who learned the language in childhood but did not reached NS status due to a language shift before adulthood (Murphy, 2014). This definition is one of the most commonly adopted in educational research, partially due to its ease of application since HLs can be identified by pre-existing levels of proficiency upon entering the classroom. In their scoping review of 34 empirical Arabic heritage studies, Visonà and Plonsky (2019) identified that 41% of

⁶ These tasks were taken from the previous specification, which, whilst still precipitating these results, was supposedly easier, according to the views expressed by teachers in conversation with the researcher at conferences.

these studies adopted the proficiency-based definition provided by Valdés (2001), which identifies a HL as someone who had been exposed to the language in their home environment at an early age and who had reached some degree of bilingualism. In his chapter providing a detailed overview of HL research in Arabic, Albirini (2016) adopts this definitional category. He identifies as HLs of Arabic ‘children of immigrants who grew up speaking their parents’ language at home and then shifted to a dominant language as they grew older because of increased contact with speakers of the dominant language’ (p.295).

Ancestral-based definitions, on the other hand, categorises learners who have a personal, historical, and/or ethno-linguistic connection to the language as HLs. These learners may not have had any previous exposure to the language, putting them potentially, in terms of proficiency, in the same category as *ab initio* learners. According to Comanaru and Noels (2009), such learners may have historical ties to a country where the language is spoken or have been raised in a heritage language-speaking community abroad. This broadened definition has some appeal to researchers of Arabic, given the religious connections between Arabic and Islam, as a result of which Muslims are often raised within families or communities where the language is used for ritual purposes, including the reading of the Quran (Scott-Baumann & Contractor, 2012). Though distinguishing between *ab initio* learners and ancestral heritage learners, such as those exposed to the language through their faith, may not necessarily prove useful in determining proficiency differences, the motivations of such categories of students differ.

A small number of studies have attempted to establish a specific definition of ‘Arabic HLLs’ under the ‘ancestral’ umbrella. Husseinali (2006) identifies three distinct groups of learners in his study of Arabic learners in an American university. The first group includes learners of Arab descent, the second non-Arab Muslims, and the third students of all other ethnic and cultural backgrounds. He goes on, however, to collapse the first two groups into one, identifying the only criterion for qualifying as a HLL, therefore, as an affinity to the Arabic language. He reported significant differences between the two resulting groups in terms of their motivational orientations, on the basis of which he proposes that they be taught in separate classrooms with tailored approaches. Mango

(2011) also proposes two groups of HLs: those who come from homes where Arabic is spoken, and those whose connection to the language is predominantly through religion. Little theoretical basis for this definition, however, is offered, although it is in line with the proficiency/ancestral distinction hitherto outlined. In their investigation of the profiles of study abroad students in Cairo, Ibrahim and Allam (2006) propose perhaps the most detailed breakdown of HLL categories available thus far in the literature. They divide learners into four categories: 1) those with Arabic as a home language, 2) those with one Arab parent, but where Arabic is not spoken at home, 3) non-Arab Muslims, and 4) Arabs who lived in Arab countries but who attended international educational institutions.

For each category, it is important to note that the nature of the language HLs are studying versus what they have been exposed to may be different. Conflating the varieties of Arabic—the modern standard generally taught in classrooms, the classical variety used in religious practise, and the colloquial used in Arabic-speaking homes and communities—runs the risk of overlooking significant differences in learner profiles. This diglossic element has been largely neglected in HL research (Albirini, 2016), and needs to be addressed in future studies.

2.2.5. Arabic Motivation Studies

In an extensive search for motivation studies focussed on L2 Arabic and published in English after 2,000, the researcher found 28 studies in total (see Appendix A.2 for details). Results were generated based on searches for the terms ‘Arabic language motivation’ and ‘Arabic L2MSS’, although the 28 results are not confined to L2MSS studies. Of the total, four were dissertations (including the researcher’s own MSc dissertation), and the remaining 24 were journal articles. While the majority of studies included motivation as their primary focus, usually in conjunction with other variables, a small number dealt with the topic through the prism of other research lenses, such as heritage learners, and attitudes and beliefs. In a reflection of the general trend in motivation studies, 25% of the studies employed mixed-methods designs (although in some cases this referred only to the inclusion of open-ended questions in their survey instrument), with a further three adopting qualitative methods. Questionnaires were the

prevailing research instrument, coupled with interviews in the majority of the mixed-methods papers.

Of the 28 studies, 16 were conducted in the United States. The remaining 12 studies were dispersed over other countries, including Malaysia, Indonesia, Iran, Nigeria, Libya, Scandinavia, Israel, South Korea, UAE, Saudi Arabia, and the researcher's own earlier study which was conducted in the UK. Nine studies were conducted with school-aged pupils.

The studies adopt an eclectic mix of theories, though for at least a quarter no methodological framework for operationalising motivation was provided. In the majority of these cases, it appears that Gardner's (1985) work, often loosely, guided an interpretation and discussion of results since the terms 'instrumental' and 'integrative' were used. When this particular research study was commenced in 2015, no studies focussing on the L2MSS in an L2 Arabic context could be identified. This recent search however, proffered seven studies adopting the model in their research design.

The studies present a diverse, sometimes contradictory array of findings regarding motivation in general. Using Gardner's AMTB, Ajape, Mamat, and Abdul Azeez (2015) studying university students in Nigeria reported high levels of integrativeness amongst learners of Arabic and low levels of instrumentality. In Saudi Arabia, Abdelhalim and Alqubayshi (2020) examine, amongst other things, the motivation of adult learners of Arabic enrolled in a university course, finding also that integrative motivations were higher than instrumental, and that overall, the level of motivation was moderate. Aladdin (2010), conversely, reported that instrumental motivations lay primarily behind Malaysian students' learning of Arabic, followed by intrinsic and instrumental motives. Brosh's (2013) findings further corroborate Aladdin's results. His participants were college students in the USA who cited 'interest in the language' as the most common reason for studying the language, followed by 'employment'. The contradictory findings of these studies are likely explained by the diverse national contexts in which they were carried out, indicating the importance of taking into account specific environmental factors.

In addition to the complexities seemingly arising from the interplay between learners and their national and educational contexts, several studies also raise pertinent points regarding learners of Arabic and the nature of their motivations for studying the language.

Ramezanzadeh (2015) conducted the first study into the GCSE Arabic student population in the UK. Although its primary focus was the performance of heritage and non-heritage learners, the qualitative component of this study, consisting of open-ended survey items and semi-structured interviews, uncovered high levels of demotivation amongst a sample of 75 students in schools around England. Of the 47 students in the sample for whom the Arabic GCSE was compulsory, almost half (46.8%) expressed a desire to drop the subject if possible. Although this is admittedly a small sample, it hints at the existence of a strong trend of demotivation amongst students of GCSE Arabic, which requires investigation.

Those who did express a desire to acquire Arabic offered a variety of reasons in response to an open-ended questionnaire item: 'What are your main reasons for studying Arabic?' These answers were coded into *instrumental* and *integrative* categories, drawn from Gardner's S-EM (2001) (Table 2.5). The main *instrumental* orientation, and the most prominent overall, was *Religious*—a desire to understand religious texts, primarily the Quran. Since all participants were from Muslim schools, this finding is not surprising. However, it is noteworthy given that the content and focus of the GCSE is MSA and does not incorporate any liturgical material. On the other hand, career-related responses such as using Arabic in as an international teacher, were almost negligible (4%). In other studies, however, such responses emerged as the most common reasons for learning a language in other studies of Arabic learners (e.g., Winke and Weger-Guntharp, 2006). This finding in Ramezanzadeh (2015) may be due to restrictions on girls' career prospects in Southeast Asian cultures (Ghuman, 2005), or a lack of awareness about particular jobs that require Arabic in contrast to other career paths. However, further investigation is required to determine the validity of these interpretations.

Integrative orientations were also low in Ramezanzadeh (2015), as illustrated in Table

2.5. Just over a third of students listed ‘communication with Arabic speakers/family’ amongst their motivations for learning Arabic. There may be population-specific reasons for this discrepancy, namely the limited opportunities perceived by students to communicate with Arab speakers given the demographics of their communities, as well as cultural restrictions on women traveling abroad alone (Mohee, 2011). Though low, the fact that ‘Communication with Arab speakers in the UK’ was the second most popular category despite the course containing no colloquial Arabic is problematic and suggests a lack of awareness concerning the nature of the Arabic language.

The study also revealed that almost 20% of students were only studying Arabic because it was compulsory in their school, giving no other personal motivation, noting in their interviews and questionnaire responses that they were taking it ‘because I have two [sic].’ A further 17.3% of students simply wrote ‘to learn a language’, or a variation of this statement, with no explanation as to what use they were hoping to put it. These responses could possibly be interpreted as students taking the subject only to fulfil the language criteria in their list of GCSE options.

Table 2.5. Orientations of Students towards Learning Arabic (Ramezanzadeh, 2015)		
Orientation	Freq.	%
Instrumental		
Religious	37	49.3
Career	3	4.0
Integrative		
Communication with UK Arabic speakers	17	22.7
Communication with family	4	5.3
Communication in Arab countries	5	6.7
Other		
Enjoyment/ interest	8	10.7
To learn a language	13	17.3
Compulsory	14	18.7

These figures represent only a cursory glance into the motivations of students towards studying the Arabic GCSE. The high level of demotivation amongst these students, along with the motivational patterns which appear to diverge from studies of motivation conducted in other learning contexts, makes this an interesting area for further

exploration. Though this study did employ qualitative data collection methods allowing for students to express their motivations in their own words rather than being constrained by a pre-existing model, the framework for interpreting these results is a limitation, as is the lack of longitudinal measures mapping the development of learners' motivations during the course. The study does, however, highlight the importance of religion to these learners as a driving force behind their desire to learn the language, as well as the wish to communicate with Arabic speakers, despite the fact that neither the colloquial nor classical is covered in the syllabus. MSA-related goals such as careers or comprehension of media are either muted or absent, suggesting a mismatch between course content and learners' motivations, all of which needs to be explored.

Other studies involving pre-university learners have been conducted in the USA. Temples (2010) carried out a study into the motivations of 80 school-level Arabic learners, whom she divides into: heritage learners (HLLs) (students of Arab descent); religious heritage learners (RHLs) (non-Arab Muslim students); and foreign language learners (FLLs) (non-Arab, non-Muslim students). Her findings reveal different motivational patterns to Ramezanzadeh (2015). The most common response to why students were studying Arabic overall was school requirement (71.3%). Religious motivations came to only 27.5% overall. Almost half claimed their parents wanted them to learn the language (45%). Only 12.5% said they were learning the language to communicate with their relatives, indicating possibly that they understood the variety they were learning was limited in how much it could help them in this endeavour.

When broken down by group, however, the figures for each motivation shift, highlighting the importance of taking students' linguistic backgrounds into account. No FLLs chose *religious* motivation, whereas over 80% of students from non-Arab Muslim backgrounds did, compared to only 30% of those from Arabic heritage backgrounds. Whilst career-driven responses were low across all groups, the highest scoring were FLLs with almost 30%, the lowest group being RHLs with just over 10%. This suggests that students from different backgrounds are motivated to learn Arabic for very different reasons. This is despite the fact that they are often integrated into the same classrooms and taught the same curricula, and that different language varieties lend themselves to the fulfilment of different language learning motives. Close attention,

therefore, needs to be paid to the demographics of students in any measurement of their motivation.

Winke and Weger-Guntharp (2006) conducted a similar study of 326 college students of Arabic in the USA, the findings of which indicate that learners' primary motivations vary according to proficiency level and L1 background. Responses to the open-ended question 'What are your primary motivations for learning Arabic?' were coded according to 10 categories established by the researchers. The most frequent response, in contrast to the findings of Ramezanzadeh (2015) and Temples (2010), was 'better employment options' (26%). 'Religious' motivations came fourth in the list (18%) and 'travel or living abroad' came sixth (15%). When the results were analysed controlling for proficiency, a significant relationship ($p=.02$) was reported between listening and the collapsed categories of *employment options* and *political/military reasons*. This suggests not only that learners have different motivations for learning the language, but that these motivations are tied to specific skills. Although no significant relationships were reported between other skill-specific proficiencies and motivations, this would be interesting to investigate in other learner contexts.

Chi-square analyses in Winke and Weger-Guntharp (2006) also revealed a significant relationship between learners' first language and different motivations. These indicated that native speakers of other languages (NSOLs) were primarily learning Arabic for religious reasons, whereas students with English L1 (NESs) were not ($p<.001$). Additionally, significantly more Arabic heritage learners (AHLs) were studying the language to communicate with family members than NSOLs ($p<.001$), but significantly more NESs were learning Arabic in order to travel or live abroad than AHLs and NSOLs ($p<.01$). This adds further support to the notion that motivations differ according to learners' demographic characteristics. Again, however, Winke and Weger-Guntharp measure motivation as a static construct, and no insight is provided as to how these motivations may develop. Furthermore, it is not clear where the 10 independently established categories for coding the data came from, suggesting again the lack of a theoretical framework.

In other contexts, other findings on learners' motivations have also emerged. Abu Bakar,

Sulaiman, and Rafaai (2010) conducted one of the few studies employing Self Determination Theory (SDT). This theory locates motivations along a spectrum spanning intrinsic (internal) and extrinsic (external), depending on the source of the motivational drive (Ryan & Deci, 2000; see Figure 2.7). In this case, the researchers applied the SDT to their examination of motivational orientations amongst 228 Muslim learners of a compulsory Arabic course in Malaysia. They used a 45-item, 7-point Likert-scale questionnaire, from which 28 items emerged as statistically reliable. Their original items spanned the spectrum of motivational orientations covered by the SDT. Of these, only 4 (*Intrinsic Motivation-Knowledge*, *Intrinsic Motivation-Accomplishment*, *Extrinsic-Motivation Identified Regulation*, and *Amotivation*) emerged as reliable and valid, with *Intrinsic-Knowledge* accounting for 34% of the variance of the outcome measures.

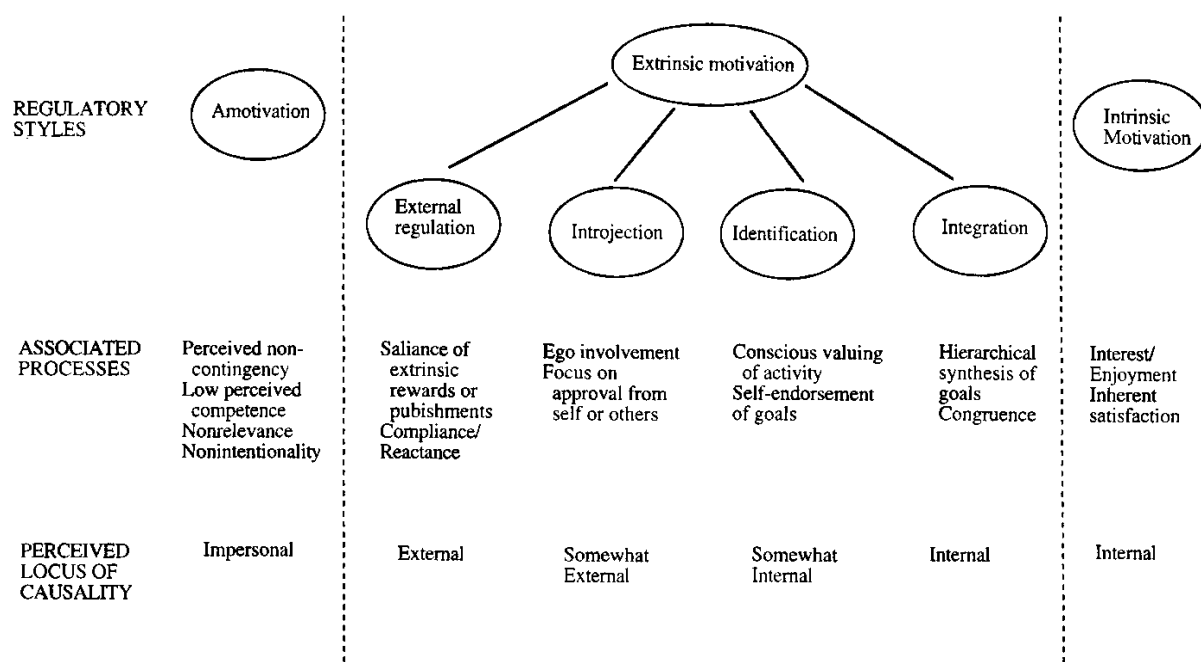


Figure 2.7. 'A taxonomy of human motivation' (Ryan & Deci, 2000)

In addition to these factors, another emerges consisting of four items, which the researchers label 'Religious Motivation'. This explains a small but significant percentage (5.43%) of the variance. This consists of items with both *intrinsic* and *extrinsic a priori* labels. The researchers claim that items with 'religious connotations' constituted a 'separate latent component of motivation in learning Arabic and could no longer be expected to adhere to their originally intended

subscales' (p.79). Whilst the emergence of this factor could be claimed to demonstrate the strength of religious motivations amongst Muslim students of Arabic, which certainly needs to be addressed in future studies, the conclusion that this should be treated as a separate subcomponent of the SDT is questionable. The existence of other items with 'religious connotations' appearing in other subcomponents (e.g., 'To show myself that I am a good Muslim'/ 'To speak the language of the Prophet') undermines the strength of this argument and suggests that these items were tapping into a different construct. Furthermore, there was no qualitative analysis of the nature of the origins of these motivational orientations. Labelling them as 'religious' motivations appears to situate them outside the intrinsic-extrinsic spectrum of motivations. Whilst it may be that religious motivations span this spectrum, it is likely that the sources and degree of internalisation of these motivations differ for each learner. For example, a person's own faith, social pressures, the promise of divine reward, the desire for spiritual improvement, among other sources could influence their interactions with the language learning process.

In Israel, Donitsa-Schmidt et al. (2004) conducted a study of the motivation of 692 elementary school students (4-6th grade) studying Arabic. This is a unique but highly interesting geo-political context given the fact that Arabic in Israel is of low prestige and is spoken by a minority. Further, in this context, negative attitudes and stereotypes of the Arabic-speaking community are pervasive. Arabic provision generally consists solely of MSA. However, children in Israel are exposed to the Palestinian dialect in their immediate surroundings. The sample consisted of students from 14 schools in Tel Aviv. Of these, nine schools taught spoken Arabic (experimental group, N=539), and the other five taught MSA (control group, N=218). Self-report questionnaires were administered, measuring students' attitudes towards Arabic, motivations to study the language, and degree of satisfaction with Arabic classes. Motivation was measured via 12 items using a 4-point Likert-scale, and satisfaction with the learning experience measure via three items with the same scale. Neither the items nor the framework for measuring motivation is provided in the research methodology section; however, reference is made in the discussion to integrative and instrumental motivations. The experimental group reported significantly more positive attitudes towards the Arabic language and its culture

($p<.001$) and higher levels of motivation overall ($p<.01$) than the control group. The best predictor of these learners' motivation was their satisfaction with their Arabic study program, although their parents' attitudes also constituted a strong predictor. This study highlights the effects of teaching different strands of the Arabic language on students' motivation, appearing to show a boost in students' experience of the learning process when taught a variety they can apply to their immediate surroundings. The relationship between motivations and language varieties therefore appears to be an essential avenue to explore in future studies.

2.2.5.1. L2MSS Studies in the Context of Arabic Learning

As aforementioned, few studies have examined the application of the L2MSS in the context of Arabic. One such study is that of Thompson (2017) who conducted a qualitative study into two learners, one of Arabic and one of Chinese, using the L2MSS and CDST as a framework. According to the researcher, 'the findings also illustrate the dynamicity of the psychological aspects of self, as these selves fluctuate throughout the language learning experience' (p.38). She observed that the 'anti-ought to self'—defined as 'reactions against external expectations' (p.42) or as she also describes it 'basking in the delight of doing the unexpected' (p.48)—emerged as a prominent component in each learner's narratives. She also reported that there was a synergistic relationship between each of the self-concepts. The anti-ought to self, Thompson argues, adds a valuable dimension to the L2MSS by encapsulating learners' 'rebellious' self-components. While the study highlights a potential addition to the L2MSS, further study is needed to determine the generalizability of this construct to other learning contexts. This being said, the in-depth CDST approach affords the researcher the ability to identify context specific variables and examine their interactions with other existing factors at play *without* the need for these findings to apply across contexts.

Another study which applies the L2MSS in the context of Arabic learners is Kong, Han, Kim et al. (2018). These researchers compared the motivations of college-level students (N=1296) in South Korea studying commonly taught languages (CTL) and less-commonly taught languages (LCTL) (either Spanish, or Arabic (N=401)). Using

multi-level structural equation model (SEM) analysis, the researchers reported that, in both cases, the IS had a stronger impact on effort than the OTS. They also measured attitudes, international posture, and competitiveness, all of which interact differently with the self-components for CTLs and LCTLs. Whilst this study sheds some light on the relationship between the self-concept components and other variables, a limitation of the study is that the researchers do not distinguish between the profiles of Arabic and Spanish learners, limiting the degree to which these findings can be said to be specific to Arabic.

A recent study by Calafato and Tang (2019) examined the motivations of 114 secondary school students (non-Arab expatriates attending private UK curriculum schools) simultaneously studying Arabic and English, and in some cases an additional language, in the United Arab Emirates. The findings specifically in relation to Arabic suggest that ‘participants’ motivational selves hint at a decidedly lukewarm attitude, and it appears as if they do not really see a strong reason to study it’ (p.33). Across the sample, learners’ IS display less motivational intensity for studying Arabic than for studying English. These low levels may be explained by the social contexts of the participants, who may not live or spend time with speakers of Arabic, despite living in the UAE, where it is the official language. Those who do exhibit higher levels of motivation appear to do so for education/training purposes, which the researchers claim could be explained by the fact that Arabic is a compulsory school subject.

Another recent study to apply the L2MSS to the context of Arabic learning was conducted by Blair and Azaz (2019). They examined the motivations of university learners in the USA, exploring the *own/other* distinction in the context of the OTS and investigating the role of episodic memory in relation to L2 self-concepts. In their study of 54 American university-level students of Arabic, they reported that the L2LE played a pivotal role in enhancing the effect of self-concepts on Intended Effort. When it came to the OTS, some evidence was reported for the bifurcation of this construct into OTS/*own* and OTS/*other*, since these displayed opposing effects on the criterion measure, the former having a positive, and the latter a negative effect. The researchers concluded that ‘it seems that L2 Arabic learners may

become demotivated when an L2 self-guide is formed from more extrinsic and preventive measures and attempts to influence intended effort directly' (p.199). However, it must be noted that neither of the pathways from the two self-concept were reported to be significant. Pathways between the self-concepts and IE were only reported to be significant when mediated by the L2LE.

Blair and Azaz (2019) also report a novel finding concerning the role of episodic memory, 'a form of memory based on the recollection of past experiences' (p.193), which was reported to have selective associations with some L2-self-concepts. This is, tentatively, in line with Markus and Nurius's (1986) belief that future self-concepts are linked to memories of personal past experiences. Episodic memory was observed to be significantly positively correlated with both OTS concepts, and, interestingly, negatively with the IS concept. This offers provisional evidence for the expansion of the L2MSS to incorporate learners' past experiences into the motivational selves framework.

Given the small sample size, the conclusions of Blair and Azaz's (2019) study are tentative and clearly require further investigation. In addition, from the description of course contents, it appears that learners were all taking MSA, with some required to take classes in a dialect, though learners were not categorised on this basis. It is also not clear for which of the students the Arabic courses were compulsory or voluntary, highlighting the need for a finer-grained focus on the actual nature of the subject(s) being studied, which may contribute to explaining some of the unexpected results.

2.2.6. Summary

Together, the findings of the aforementioned studies present a fragmentary picture of the motivations of Arabic learners, indicating the necessity for further investigation in multiple areas. Generally, studies have employed simplistic approaches to measuring motivations amongst Arabic learners, either through open-ended questionnaire items or by providing a list of common orientations towards language learning. The majority of studies employ a rudimentary version

of Gardner's AMTB and none include longitudinal measures. Future studies, therefore, may be able to provide further insights by adopting longitudinal designs, along with qualitative measures, which mine the depths of reasons behind learners' desire to study the language.

Despite these limitations, the studies do, however, offer a number of leads which future research into motivation should pursue. Firstly, they indicate high levels of variation amongst students as to their reasons for studying the language. This is seen to be dependent on several factors including heritage learner status, first language, and faith. Religion and its interaction with motivation is an area which requires further scrutiny, in particular through qualitative study, in order to explore the relationship between religious beliefs and the learning of Arabic and the situating of religion along the internally-externally derived spectrum of motivation. Motivation also appears to be affected by language variety, with the learning of a spoken variety promoting motivation in a context in which access to Arabic-speaking communities is available. The complexity of Arabic usage in authentic contexts and the perceived difficulty of the language, present a number of theoretical challenges to learners' motivations, both their initial desire to learn the language, and their drive to continue once they have commenced their study. Despite the need for research into the motivation of AFL learners given its growing importance pedagogically and politically, it has been limited particularly in the UK. Attention should be paid to the nature of classroom content and skills being developed, and how these affect learners' motivations to study the language. In the studies which employ the L2MSS, researchers tend to integrate other measures of individual differences, such as international posture and metacognition. Whilst offering some evidence for the validity of this framework in the context of L2 Arabic, they also point to the complexity of the construct and the plethora of interacting variables, necessitating a nuanced and detailed approach to examining the interplay between course content, learner, and context in order to truly grasp the dynamics at work in the realms of Arabic GCSE learners' motivations.

3. METHODOLOGY

3.1. Research Overview

This research examines the motivations of Arabic language learners in UK secondary schools through the lens of the Second Language Motivational Self System (L2MSS). This involves the development and testing of new instruments designed to measure learners' motivation, conceptualised as the strength of, and discrepancy between, their Arabic-specific current and future self-concepts. The study also measures proficiency, both through purposely developed tests and self-reports. The study aims to explore the validity and nature of, as well as interactions between, the self-concepts and proficiency in the context of this multiglossic language via use of a longitudinal mixed-methods design.

3.1.1. Rationale for Adopting a Mixed-Methods Approach

Mixed-methods designs are growing in prominence in the field of motivation studies, following a period of almost exclusive focus on quantitative approaches (Al-Hoorie & MacIntyre, 2019; Boo et al., 2015). Mixed-methods designs combine quantitative and qualitative methods, treating them not as oppositional in terms of their paradigmatic stances, but as compatible and complementary. The respective data elicitation techniques are seen to enhance the conclusions drawn from each other, enabling the researcher to present robust and detailed findings. This 'methodological eclecticism in answering research questions' (Hessel, 2016, p.69) allows the researcher to draw from a wide array of possible data collection methods, using techniques most appropriate to the research questions and expertise of the researcher (Tashakkori & Teddlie, 2010).

The rationale for adopting a mixed-methods design approach in the present study stems from theoretical and experiential motivations. Firstly, this method sits in line with the researcher's own adherence to a pragmatist paradigm. Pragmatism 'holds that human actions can never be separated from the past experiences and from the beliefs that have originated from those experiences. Human thoughts are thus intrinsically linked to action' (Kaushik & Walsh, 2019). From a realist ontological perspective, adopted in this

work, mental states are regarded as ‘real’ phenomena which interact with other more ‘concrete’, observable variables. This makes capturing both actions and participants’ related perspectives and beliefs intrinsically important when exploring a particular research domain. A pragmatist approach best facilitates this, since it rejects the incompatibility hypothesis (Lincoln & Guba, 1985), viewing quantitative and qualitative methods as: (a) corroboratory, whereby findings from each validate or confirm each other; (b) complementary, since multiple methods used in answering the same question can ensure wider scope and depth of analysis is achieved; and (c) developmental, in cases where one method highlights oversights in another and offers areas for expansion or improvement (Hessel, 2016).

The CDST movement, which is gaining traction in the field of motivation studies, posits that individuals and constructs form part of, and contain within themselves, complex systems, with reciprocal, rather than solely linear inter-relationships in operation (Dörnyei & Ryan, 2015). In a study which attempts to interrogate the validity and nature of constructs in a new language learning context, and at the same time investigate the dynamics of these constructs at a granular level, a mixed-methods design offers a way to simultaneously capture an image of the system—at least in part—and then zoom in on individual components either in isolation or in relation to other system components. Originating with Campbell and Fiske (1959), the triangulation of data, which describes the combination of different methods in the examination of a single phenomenon, lies at the heart of a mixed-methods approach. This aids in the capturing of this complexity by combining multiple perspectives, thus allowing for the construction of a more multidimensional image of the system.

Finally, an underlying aim of this research was to push open the door onto the world of Arabic learning in the UK. Some headway was made in this regard through the researcher’s own Master’s study (Ramezanzadeh, 2015), which also employed a combination of quantitative and qualitative methods. It was this approach that led to the current investigation of motivation in this research. As a result of shining multiple lights on a subject, not only are answers to the research questions revealed with greater clarity, but the breadth of the beam is wider, thus highlighting multiple areas for future

study, all of which are desperately needed given the sparsity of quality Arabic-related research.

3.1.2. Researcher Positionality

In the interests of determining, at least to some degree, how the researcher contributes to the construction of the reality being examined, reflections on the researcher as an ‘observation instrument’ are called for.

The researcher is a woman of mixed heritage, who wears a headscarf which makes her identifiable as a Muslim. In the context of some of the schools in the sample, this affords her a degree of ‘insider’ status, although this was somewhat counteracted by her ‘foreign sounding’ name and ethnic ambiguity. As a non-native Arabic speaker, who studied the GCSE herself, she was able to invoke a shared experience with the students in her introduction to the research, which it was anticipated would alleviate some of the potential hesitation learners may have felt speaking to an ‘outsider’, a researcher from the University of Oxford.

Her interactions with the research study are informed, to varying degree, by her experience: as a foreign language learner of Arabic, who sat the GCSE and A-Level in Arabic while at secondary school and studied Arabic at university during which she lived in the Middle East for a year; as a member of staff at two Islamic schools; as an Arabic teacher at both school and university level; and through interacting with the exam board, having been asked to review the new specifications and exam materials prior to their adoption nation-wide in 2017.

This combination of factors—personal and experiential—positioned the researcher both as an ‘insider’ and an ‘outsider’, affording her unique perspectives into this field of study (Taylor, Bogdan & DeVault, 2016). Her awareness of and consideration of these factors prior to and during the research was a constant feature of her interaction with schools and students. No researcher is an entirely objective observer, but in identifying key components of the perspectives they bring to the subject of study, researchers can identify and be transparent about the inbuilt lens through which they experience their

particular field. Reference is made to these features in subsequent sections when apparent in analysis or interpretation of the data and findings.

3.1.3. The Second Language Motivational Self System (L2MSS)

Motivation is conceptualised as a combination of both the impetus and the drive to continue learning; in other words, the direction in which a learner is aiming and the intensity with which they pursue this goal.

To operationalise this concept, this study adopts and develops Dörnyei's (2005) L2MSS model, measuring the motivations of secondary school learners of Arabic taking the General Certificate of Secondary Education (GCSE) examinations in the subject. Dörnyei's original L2MSS model of motivation incorporates the *Ideal Self* (IS), the *Ought-To Self* (OTS) concepts, and the *L2 Learning Experience* (L2LE), as illustrated in Figure 3.1.

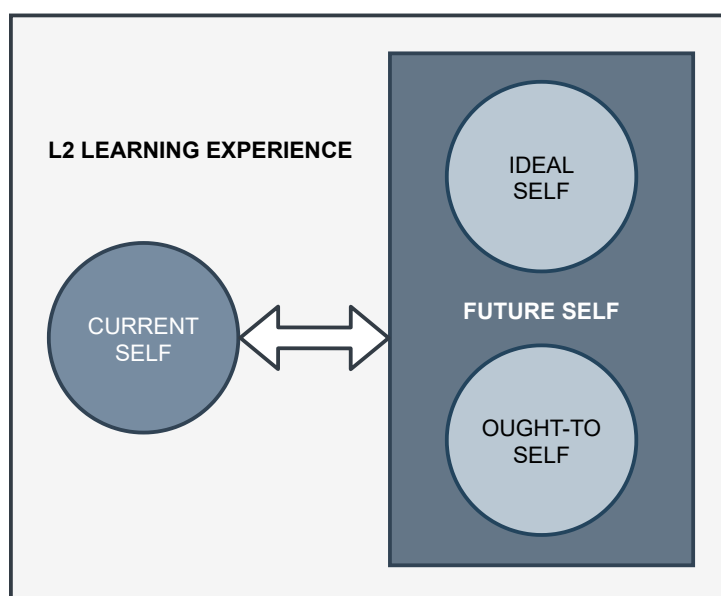


Figure 3.1. The Second Language Motivational Self System

The *Ideal Self* (IS) refers the L2-specific aspect of a learner's future self-concept (Dörnyei, 2009, p.29); the *Ought-to Self* (OTS) concerns the future self a learner believes they should achieve in relation to the L2 to 'avoid possible negative outcomes' (Dörnyei,

2009, p.29). The *L2 Learning Experience (L2LE)* is a broad category encompassing both the learning environment and learners’ interactions with it.⁷

Dörnyei originally drew upon the work of Markus and Nurius (1986) and Higgins (1987) to define motivation as the strength of a learner’s domain-specific current and future self-concepts and the desire to reduce the discrepancy between them. This is the definition adopted in this research. Measurement of learners’ IS and OTS concepts are taken along with, critically, an operationalisation and measurement of their *Current Self (CS)*-concept, which has been largely neglected in L2MSS research. These three self-concepts are measured across three different varieties of Arabic: *Classical*, *Modern Standard*, and *Colloquial* varieties. These three varieties represent the specific domains of the self-concepts measured. The study develops Hessel’s (2015) survey instrument to measure each self-concept in relation to the three Arabic language varieties (as illustrated in the matrix in Figure 3.2). Three sets of bipolar scale items each under a separate language variety-related opening statement were developed and utilised to generate and examine scores for learners’ motivational self-concepts at two timepoints during the school year.

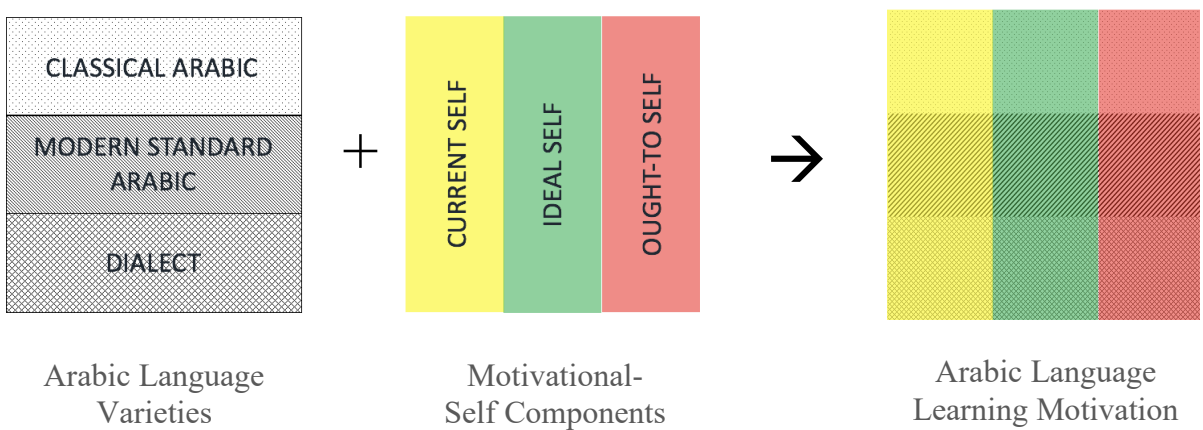


Figure 3.2. Arabic Multiglossic Motivation Matrix

⁷ Dörnyei (2019a) has most recently suggested that the L2LE could be conceptualised as: ‘the perceived quality of the learners’ engagement with various aspects of the language learning process’ (p.19). In the original design of this study, the L2LE was examined through the prism of engagement. Unfortunately, participant non-compliance meant the data collected for this dimension did not meet the criteria for inclusion.

3.1.4. Proficiency Measures

According to Al-Hoorie (2018), few studies have examined the relationship between self-concepts and objective measures of learning outcomes, such as achievement or proficiency. To address this, unique proficiency tests specific to the target population of GCSE Arabic learners were formulated, examined, and validated to provide accurate measures of learners' proficiency. These consist of a Yes/No or 'Checklist' test and a cloze test, both of which have been used extensively to measure proficiency in other languages, though not in Arabic. Learners' Arabic language proficiency is explored in conjunction with their motivation to explore the relationship between these multiple constructs.

Learners' self-rated proficiency scores across the three domains of the language are also captured to assess and compare these with proficiency test results and determine whether there are any differences in the interactions between motivational selves and the subjective and objective proficiency measures.

The different self-concepts are posited as interacting with each other over time, as well as with proficiency both at and between their respective timepoints (see Figure 3.3). As well as an analysis of the validity and nature of each construct and the quality of the measurement instruments developed, the relationship between motivation and proficiency is also explored. The effects of gender and heritage language learner (HLL) status as a variable in these interactions are also examined. Each of these relationships is explored through statistical analytical procedures, including factor analysis and path modelling.

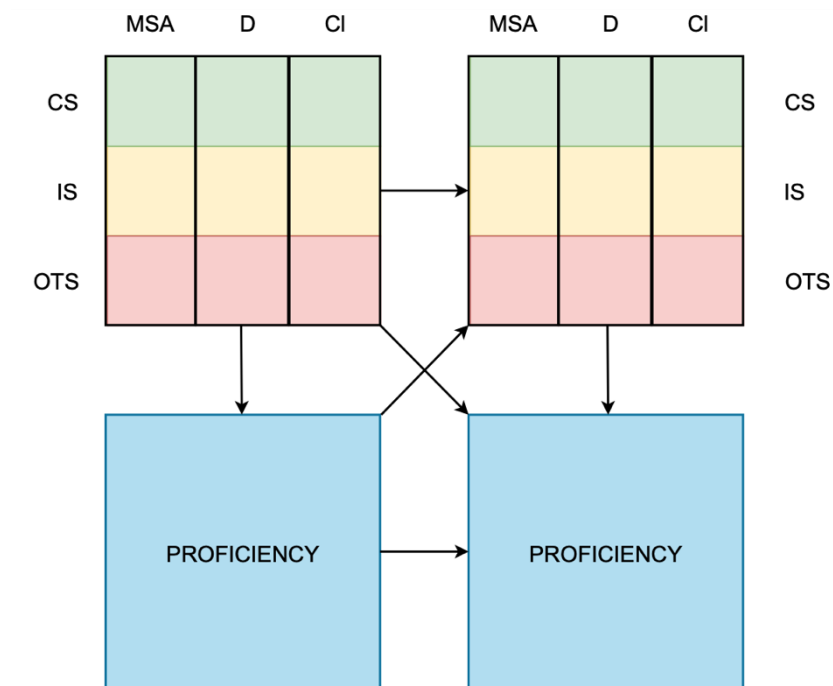


*Figure 3.3. Theoretical model of L2MSS construct interactions
(MSA=Modern Standard Arabic; D=Dialect; CI=Classical; CS=Current Self; IS=Ideal Self;
OTS=Ought-to Self)*

Motivation and its related variables are conceptualised as a complex dynamic system in which multiple interactions take place within and between constructs (Dörnyei & Ryan, 2015). To better capture this complexity, and to capture aspects of the L2LE within which learning takes place more specifically, semi-structured interviews were conducted with participants. Learners were asked about their motivational self-concepts and their learning experiences. Data from both sets of analyses are integrated and co-analysed to present as full a picture of the dynamics at play as possible.

3.2. Research Questions

In order to address the research aims of examining the validity of the measurement instruments and constructs, and to analyse the nature of and relationships between the variables examined, the following research questions were formulated:

1. What is the structural validity of:
 - a. the L2MSS construct as it relates to Arabic, as measured by the motivation survey scales?
 - b. the multiglossic self-rated proficiency construct, as measured by the self-rated proficiency scales?
2. What is the nature of Arabic learners' motivational self-concepts? How do learners' motivational self-concepts vary based on:
 - a. Arabic language variety?
 - b. Gender?
 - c. Heritage Learner status?
3. How do learners' motivational self-concepts develop over time?
4. What is the relationship between proficiency and self-concepts
 - a. Within timepoints?
 - b. Between timepoints?
5. What role does self-discrepancy play in the L2MSS?

3.3. Research Design

To answer these research questions, a mixed-methods design was formulated and executed. This incorporated cross-sectional and longitudinal data collection phases, illustrated in Figure 3.4.

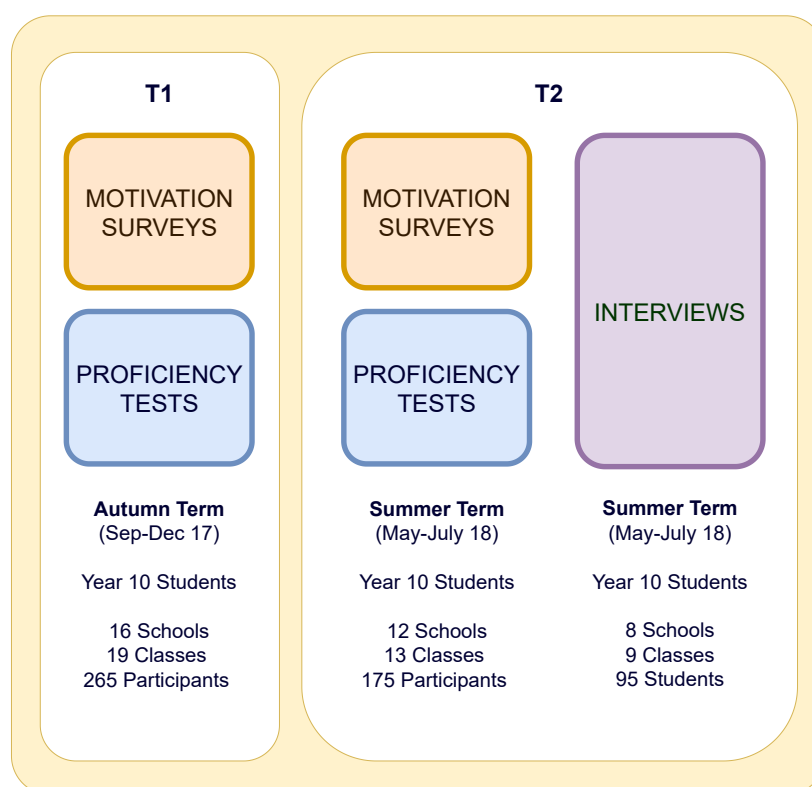


Figure 3.4. Research Design and Sample Statistics

The data was collected over the course of one academic year (from September 2017—July 2018) from 265 Year 10 GCSE Arabic language learners. Responses were first analysed in order to assess the measurement quality of the instruments, and to examine the structural validity of these motivational-self components in the context of three different varieties of Arabic: *Modern Standard*; *Classical*; and *Dialect*. Subsequently, the motivational self-concepts of these learners are explored. Specific focus is placed on how these self-concepts vary between Arabic language varieties, and on differences between genders and learners of heritage and non-heritage language backgrounds. The relationship between learners' motivational self-concepts and their proficiency is examined both within and across time-points. These quantitative findings are triangulated with interview data collected from 95 participants. Both datasets were first examined independently with the findings from each distilled. Using the 'Following the Thread' method (Moran-Ellis, Alexander, Cronin et al., 2006), the findings from each were then used to guide subsequent analysis into the other dataset, thus integrating the two in order to provide greater insight into the complex emergent patterns and relationships within and between constructs.

3.3.1. Recruitment & Sampling

To identify the population of GCSE Arabic learners, an extensive search was conducted in order to pinpoint schools offering the subject across England. The parameters for the population were set to the following: secondary school students (Year 10) engaged in studying the language as either a compulsory or optional GCSE subject in their respective schools around England. This population was chosen for a number of reasons. The Arabic GCSE, offered by the Pearson Edexcel exam board, is the only standardised examination available at this level of compulsory education. All schools⁸ were therefore studying the same content at the same level. The study focused on Year 10 students. It was decided that this group, who generally begin the study of GCSE material in this school year, would be recruited since they faced fewer exam and coursework deadlines than their Year 11 counterparts. Only schools offering classroom teaching of the subject—as either a compulsory or optional element of their curriculum—were approached. This ensured students were all of comparable ages and were receiving classroom tuition, as opposed to being entered as individual or external GCSE candidates whose learning was entirely their own responsibility. Schools in England were chosen to avoid the complication of introducing alternative examination systems present in other parts of the UK.

The results of this search were entered into a database. According to the information available, 144 schools fit these criteria at the time of investigation. The majority of these were independent Muslim faith schools. This extensive survey of the population and the subsequent database that was constructed constitute a major contribution to the field.⁹ A breakdown of the population is available in Appendix B.1.

Convenience sampling methods were employed during the recruitment of participants. Each of the 144 schools in the database was contacted by phone and email with an invitation to participate in the study. 16 schools agreed to participate in the first phase (T1). In total, 265 Year 10 students studying for the Arabic GCSE completed the first

⁸ Bar one at which they were studying the iGCSE offered by CIE.

⁹ A subsequent project by the British Academy has attempted a similar effort (see Section 2.2.3), though this was not available during the recruitment period for this study.

Motivation Survey and Proficiency Tests at the beginning of the academic year 2017. At the end of the academic year, these 16 schools were contacted again to arrange for the completion of the second batch of instruments (T2). Of these, 13 classes agreed to repeat their participation, giving a total of 175 students at T2. The researcher attended 12 of the schools at T1 to collect the data in person. Data from the remaining 4 was collected by a supervising teacher and sent via post. At T2, data was collected from 3 of these schools by post, and from 9 in person. Learners at the schools who completed the surveys and proficiency tests with the researcher present at T2 were also invited to participate in the semi-structured group interviews. Consent was obtained from 95 students of the 175 at T2 for this phase of the research. Interviews were recorded by the researcher and transcribed by a data entry assistant (DEA) for subsequent analysis.

3.3.2. Ethical Considerations

Ethical clearance was obtained from the Central University Research Ethics Committee (CUREC) of the University of Oxford after approval of the research design was received and before data collection commenced.

3.3.2.1. Consent

Schools willing to participate in the research were asked to distribute Information Sheets and Parental and Student Consent Forms to be signed and returned to the school indicating confirmation of consent. For the Motivation Surveys, Proficiency Tests, and Engagement Surveys, Opt-Out consent forms were chosen since only written data was collected. Only students for whom consent was not granted were required to return the forms. For the interviews, Opt-In versions were distributed since audio data was being recorded which requires receipt of written consent and parental agreement since participants were under the age of 16.

Students' right to withdraw themselves from the study at any point without any negative consequences was highlighted before each stage of the data collection, and an emphasis was placed on the anonymity and confidentiality of all data collected. Participants were reassured both in writing and verbally that their responses to the

surveys, opinions expressed in the interviews, and their performance on the proficiency tests would not affect their GCSE or class grades.

3.3.2.2. Impact

Though the Motivation Surveys and Proficiency Tests were designed to be completed as quickly and easily as possible, at each time-point the procedure usually took up one period of Arabic class time (45 minutes to an hour). Since Key Stage 4 is a particularly demanding time in terms of workload for both teachers and students, a great deal of flexibility was offered both in regard to timing and method of completion. Teachers were invited to stipulate a suitable time for the completion of these activities, so as to plan their delivery around them and minimise the impact of any missed lesson time. The sample was purposefully drawn from the Year 10 student population, rather than Year 11s, to avoid taking up any class time in the immediate lead up to examinations. Every effort was made to ensure that the surveys, tests and interviews were completed promptly and efficiently so as to avoid placing any extra stress on teachers and students. The researcher took extra precautions to ensure visits were as smooth as possible: always arriving early to the school, staying later in case teachers preferred to talk after school hours, accommodating last minute changes to scheduled visits.

Though less conventional, the positive impacts of the study and indeed researcher presence and contact with the participating schools and students merit mention. There is evidence to suggest that act of reflecting on motivation and future selves, whilst potentially an interfering factor from a purely methodological perspective, may have positive benefits for participants' progress (McMillan & Hearn, 2008). Though a one-off imagination of an ideal self is not enough to maintain high levels of motivation, the exploration of these concepts with the researcher during the interviews, it is hoped, will inspire learners to reflect on these future self-images in their own time. In the majority of schools, teachers expressed that they had not explored learners' future self-images with them. The indirect suggestion of the importance of this practice through the interview questions at the end of the research will hopefully also encourage some teachers to adopt this method of motivating students.

Particularly within the Muslim girls' schools, a number of teachers expressed that having an educated, identifiably Muslim woman from Oxford University interacting with them was of great value, and they were keen to increase their students' exposure to positive, relatable role-models. Several Arabic teachers emphasised to their students when introducing the researcher that she was not a native-speaker of Arabic and yet was teaching the language at a leading institution, appealing to this experience to make a career or at least further study of Arabic seem more achievable.

In the remaining time that was left between completion of both booklets and the end of the period, the researcher made a point of inviting questions from the students. Teachers in each class encouraged learners to ask questions about Arabic, studying it as a degree subject, the university application procedure, university life and the experience of being a Muslim at university. Often these questions were raised unprompted, in particular the question of how the researcher got into Oxford. Since a number of schools in the sample were situated in areas of low socio-economic status, emphasis was laid on the researcher's shared experience with having gone to a struggling school in a multi-cultural, largely underprivileged area. Students were perceived to find this narrative entertaining and inspiring. These question-and-answer sessions appeared to impact their perceptions of both the possibility of pursuing their potential, and the prospect of challenging the socio-cultural barriers to an ostensibly inaccessible institution.

In addition to the researcher's use of her own experiences and identities to give back to participants, further contributions were also made in terms of time, effort, and expertise. The researcher accepted requests to take part in a careers event; agreed to be interviewed on the current state of the education system for a school newspaper; gave an impromptu session on ways to establish and maintain motivation; delivered a poetry workshop; used personal contacts to make links between schools and the Oxford University Islamic Society; linked teachers up with study abroad programs for the purpose of school visits; offered advice to teachers on textbooks and learning materials when requested. None of these contributions were offered in exchange for participation. However, taking part in the research provided a rare opportunity for schools to benefit from links with an academic in Education and her expertise and experience in the field.

Given the isolated nature of Arabic teaching, and the struggles expressed by teachers and learners with the GCSE, the researcher was very willing to contribute where professionally possible.

Finally, concomitant to the abovementioned point, the presence of the researcher in the Arabic classroom also allowed teachers the opportunity to speak to an external party regarding their struggles and frustrations with teaching the Arabic GCSE course. It was, of course, made clear that any anticipated impact this research may have upon the provision and examination of Arabic would only emerge in the long term. Regardless, the perception of the researcher as a potential bridge between the worlds of teaching and policy making, it seems, provided an opportunity for teachers to off-load their concerns and complaints with the current state of affairs. This function of the study—a pressure valve for the building grievances of Arabic teachers—is an unexpected, though welcomed side-effect of the research, although it raises serious questions concerning the need for a continued support system for these teaching professionals, and the need for recourse to channels via which they can express their dissatisfactions and have these heard.

3.3.2.3. Sensitivity

The researcher has extensive Arabic teaching experience and has taught the Arabic GCSE at a secondary school. The researcher also benefits from having used similar data collection methods (surveys, tests, and interviews) with participants from the same population for her MSc study. She was therefore sensitive and respectful when interacting with teachers and students. She has been DBS certified and therefore legally permitted to work with children and young adults.

With this training and experience, the researcher was able to be sensitive to the cultural and religious practices of staff and students. Her work with teenagers in a variety of settings allowed her also to be sensitive to the stresses and pressures they may be experiencing within and outside of school. Particular care was taken to avoid the portrayal of any judgement, which may have been expected from an adult of shared faith or cultural background.

The research did involve potential discussion of family relations, personal and cultural expectations, and possible consequences in the absence of achievement. The researcher was acutely aware that these can be sensitive topics for some individuals. Phrases such as 'failure' and 'punishment' were avoided so as to try and avoid triggering emotions or fears that may affect participants' wellbeing. The researcher is well versed in how to access schools' safeguarding policies and was prepared to contact the appropriate member of staff in the instance of any concerns for a child's safety or wellbeing. As it happened, the need for this did not arise during the course of the research.

3.3.2.4. Data Protection

Care was taken to ensure the anonymity, confidentiality and security of the data at all times. The researcher travelled to and from each school using her own vehicle, thus enabling her to safely transport the data without fear of loss, damage, or interference. All paper-based data was kept securely in a locked office on University premises, where they were stored until such time as they had been converted to electronic format and data analysis completed. The papers will be destroyed upon completion of this study. Electronic data was anonymised and stored securely on password protected laptop. Data collected digitally (e.g., digital audio-recordings of interviews) was transferred immediately from the portable recording devices to the researcher's password protected, personal computer. The employment of a data entry assistant (DEA) was subject to an agreement on the confidential nature of the data. Audio recordings and booklets were transferred from the researcher to the DEA in stages in person. These were kept in a secure, personal workspace during digitization. Digitised data was transferred via a secure online document transfer system or the researcher's personal USB device. Upon completion of the process, all files were deleted from the DEA's computer. All identifying information relating to participants has been removed in the writing of this document.

3.4. Methods

This study examines the structural validity of the motivational self-concepts as they relate to different Arabic varieties. Since new instruments were developed in order to capture learners' motivational self-concepts and proficiency, the validity and reliability of these instruments is investigated. The survey integrates a new component, the CS, as well as fine-grained measures of the IS and OTS, hence the need for investigation to determine their accuracy of measurement. The nature of learners' motivational self-concepts and how these vary based on language variety, gender, and heritage status, and how these develop over time is next explored. Also examined are learners' proficiency levels (both self-rated and proficiency test-based) and the relationship between these and participants' motivational profiles. Longitudinal developments of learners' motivational self-concepts and their interactions with proficiency over time are studied. The measurement and role of self-discrepancy is investigated via analysis of the survey data. These self-concepts are also captured and investigated via data collected from semi-structured interviews, which also shed light on the nature of participants' learning experiences. Following the individual analysis of each dataset, the quantitative and qualitative data was then co-analysed and threads followed from one through to the other in order to fully capture the complexity of the constructs and their interactions.

3.4.1. Instrument Development

Three instruments—Motivation Surveys, Proficiency Tests, and an Interview Protocol—were designed for the purpose of addressing the aforementioned research questions. The surveys and proficiency tests were administered at two timepoints, at the start and end of the academic year. Interviews were conducted with participants following the second phase of quantitative data collection.

3.4.1.1. Motivation Survey

The Motivation Survey consists of four categories of items totalling 27 questions. These categories are interwoven throughout the instrument, as outlined in Table 3.1, in order

to avoid fatigue effects from completing multiple similarly phrased questions. The full survey can be found in Appendix B.2.

Table 3.1. Motivation Survey Components			
Section	Topics	Quantitative	Qualitative
Demographic/ background information	Ethnicity Languages spoken Religion Years and nature of Arabic study Exposure to Arabic at home/ in community	6, 7, 8, 19	
Motivational Self- concepts	Likert-scale items: - Religious texts - Fluent communication - Understanding modern written texts Reasons for studying the GCSE Future use of Arabic	9, 15, 26	23
Experiences with, and attitudes towards the GCSE	Feelings towards Arabic Feelings towards the GCSE	1, 2, 3, 4, 10, 11, 12, 13, 14, 16, 17, 18, 24	20, 25, 27
Self-rated proficiency	Likert-scale item: - Self-rated skill proficiency Grade comparison	21, 22	

3.4.1.1.1. *Demographic details*

Learners were asked to indicate their ethnic background, religious affiliations, and languages they spoke at home. The ethnicity scale was adapted from the 2011 census. The researcher's version asked students to qualify their country(ies) of origin if they ticked 'Arab', or 'Mixed' background.

Care was taken to elicit as much information as possible concerning participants' experience with and exposure to Arabic outside of school. Previous findings indicate that identifying a student as being of 'non-Arab' ethnicity was not sufficient in classifying as a non-heritage language learner, since some students had had exposure to Arabic from living abroad, or because their parents spoken Arabic as a second/foreign language at home (Ramezanzadeh, 2015). Several questions were therefore included to cover exposure to/usage of Arabic at home, in the local community, and via the internet.

3.4.1.1.2. *Motivational Self-concepts*

Measuring a learner's motivation must involve pinpointing their current and future self-concepts in order to calculate the distance between them. Based on Dörnyei's (2009) conceptualisation of the IS consisting of multiple contributing components, it is logical to infer that the CS and OTS are also multi-componential constructs. According to Dörnyei, each component of the IS affects the motivational capacity of a person's overall future self-concept. He states that a future self-image must be:

- ✓ 'Available to the learner,
- ✓ Different from the current self,
- ✓ Elaborate and vivid,
- ✓ Plausible,
- ✓ In harmony with the learner's social environment,
- ✓ Should not be regarded as comfortably within reach,
- ✓ Regularly activated,
- ✓ Accompanied by procedural strategies, and
- ✓ Offset by a counteracting feared possible self.' (p.20)

Few studies have captured, systematically, any of the separate components of the IS as conceptualised by Dörnyei and his L2MSS framework in a way that constructs the learners' self-concepts multi-dimensionally. One study which does adopt this approach was conducted by Hessel (2015). Her survey offers the clearest and most detailed operationalisation of the IS to date (see Section 2.1.2.1.2.1). The current study modifies and expands upon Hessel's design in two principal ways. Firstly, multiple scales for the CS and OTS are incorporated into the instrument, and secondly, each of these three sets of scale items are replicated under three unique opening statements related to a specific language variety and set of skills, to thus capture self-concepts across language domains.

3.4.1.1.2.1. *Item development*

Three bi-polar matrix scale questions (items 9, 15, 26) were developed. Each item begins with an opening statement (Table 3.2) illustrating a L2 future self-image, one for

each of the Arabic language varieties. These were generated by focusing on the most commonly identified domains of that particular strand of Arabic based on sociolinguistic studies of the language (Wahba, 2006).

Table 3.2. Motivational Self-Concept Scale Opening Statements		
Item	Language Variety	Opening Statement
26	Modern Standard Arabic	In the future, I will be someone who can use Arabic to fully understand the news, books, or magazines.
9	Classical Arabic	In the future, I will be someone who can fully understand religious texts in Arabic.
15	Dialect/ Colloquial Arabic	In the future, I will be someone who can communicate fluently with Arabic speakers.

Each of these are followed by 11 sub-items, divided into three factors (outlined in Table 3.3): the central concept of the IS; the OTS; and CS.

Table 3.3. Motivational Self-concept Scale Components			
Construct	Sub-component	Bi-polar scale labels	
Ideal Self	Desirability	I desire this very much	I do not desire this at all
	Plausibility	This is very likely	This is very unlikely
	Vividness	My imaginations about this are very clear and real to me	My imaginations about this are very vague
	Frequency	I imagine this very often	I never imagine this
Ought-to Self	Social Pressure	Other people want me to achieve this goal	No one else wants me to achieve this goal
	Fear of Consequences	There will be bad consequences if I don't achieve this	There won't be bad consequences if I don't achieve this
	Force	I feel forced to achieve this	I don't feel forced to achieve this
	Fear of Disappointing Others	People around me will be disappointed if I don't achieve this	People around me will be fine if I don't achieve this
Current Self	Self-perception	This is true of me now	This is not true of me at all right now
	Confidence	I am very confident in my ability to [...]	I am not confident in my ability to [...]
	Ability	I can fully understand [...]	I can't understand [...] at all

Ideal Self

The IS component of motivation was modelled on Hessel's instrument in which it is operationalised as consisting of four sub-components: plausibility, desirability, accessibility and self-discrepancy. These constituent components were drawn from Dörnyei's list of 'necessary conditions for self-images to energize motivation' (2015, p.92). Some of the items in this study were adapted directly from Hessel's research, others were reworked, and additional items were added.

Desirability: This has constituted the primary focus for measurements of the IS in most studies. The extent to which a certain future self-image is desired is strongly linked to its motivating potential (Markus & Nurius, 1986), and in Hessel's (2015) study this was strongly correlated with the amount of effort students were willing to make in learning the language.

Plausibility: This is conceptualised as the likelihood of attaining a specific future self-image. Whilst Hessel's study did not find a significant correlation between plausibility and effort, this component has been supported by a number of other studies, including Oyserman and James (2009), and so was retained.

Frequency: This label was used to replace the term 'Accessibility' in Hessel's study. The original condition identified by Dörnyei referred to the necessity of a future self-image as being 'regularly activated', and this is what the original wording of Hessel's item appears to measure: 'I imagine this often/ I never imagine this'. Frequency of activation has been empirically supported as a mediating feature of motivation (Norman & Aron, 2003), and also correlated with amount of effort expended (Hessel, 2015), due to the fact that levels of motivational behaviour are partially contingent on learners' capabilities in generating mental imagery (Dörnyei & Chan, 2013; Kim & Kim, 2014).

Vividness: This item was developed from the third of Dörnyei's conditions, which refers to the necessity of vividness, detail, and elaboration in the

motivating capacity of a future self-image. This item has not been tested before, though it is suggested as a potential expansion to Hessel's survey in her conclusions.

Ought-to Self

The OTS, though often theorised, has received little empirical testing when compared to the IS. According to Dörnyei (2015), 'The Ought-to L2 Self reflects the attributes that one believes one ought to possess to meet expectations and to avoid possible negative outcomes in the process [...] It may bear little resemblance to one's own hopes or desires since these represent someone else's vision for the L2 learner in question and thus they concern an 'imported' image of the future that the learner will then internalize to some extent' (p.87-88).

This definition constituted one of the primary sources for the development of the OTS items. One of the few studies to examine the OTS via the means of multiple components was Taguchi et al. (2009). They measured the OTS using six items, some of which were also adapted for this study. They also included a section on Parental Encouragement/Family Influence, some of which was integrated into the OTS section. The researcher's own knowledge of the population based on her MSc study, and her personal experience having taken and taught the Arabic GCSE were also drawn upon.

Since this study involved young learners, it was important that these items were worded as sensitively as possible and according to careful ethical considerations since they would be reflecting on external pressure, fears, and consequences. The words 'fear', 'failure', and 'expectations' were avoided, given that these students were already in a high-pressured environment in the lead up to their GCSE examinations.

Force:	This item was designed to measure the sense of obligation, or 'ought-to-ness', which participants may feel in relation to learning the target language. The phrase 'I ought' was replaced with 'I feel forced', since it was possible that younger learners may be unfamiliar with this rather antiquated wording. The idea of

obligation was considered to be vital to this particular study since many of the participants were taking the subject as a compulsory part of their religious studies curriculum.

Fear of Disappointment: This was touched upon by Taguchi et al.'s (2009) item 'If I fail to learn Arabic I'll be letting other people down', as well as a number of items included in their survey section on 'Parental Encouragement/Family Influence' through which they enquire about learners' family expectations and fears surrounding failure. Family influence emerged as a strong predictor of voluntary learning in some cultures. This item was therefore included partly to test whether or not the findings in regard to Japanese and Korean backgrounds could be extended to learners from other cultural backgrounds.

Social Pressure: Social pressure from friends, family, and people of authority was also included in Taguchi et al.'s (2009) study. Since there were clearly multiple areas from which pressure could stem, and only limited space for the items, the wording of this was kept vague. The phrase 'other people' was used, and opportunities for expansion upon whom these sources of pressure may be included in the open-ended questions, and in the interviews.

Fear of Consequences: This is alluded to in Dörnyei's definition of the OTS when he mentions the avoiding of 'possible negative outcomes'. The items in Taguchi et al.'s (2009) study all point to specific negative outcomes which may overlook other consequences learners are trying to avoid. Again, the wording here was kept vague, using only 'there will/ won't be bad consequences' to not achieving the specified future self-image, with opportunities for elaboration in qualitative sections of the research.

Current Self

In order to capture learners' current self-concepts, three CS components were measured as part of the Likert-scale items on learners' motivational selves.

Self-perception: Hessel's item on 'self-discrepancy', which states 'this is already/not true of me now', appears to measure not the nature of a learner's future self-concept, but whether they currently perceive themselves as fitting the description in the opening statement. As such, it was felt that this was more appropriately included as a measure of participants' CS. Given Higgin's (1987) insistence that self-discrepancy could not be captured via direct questioning, current-future self-discrepancy will be ascertained by calculating the distance between learners' current and future self-concepts.

Ability: Another item designed to capture learners' current self-concept, labelled 'ability', was added in order to examine whether or not the relocation of the self-discrepancy item to the CS was in fact justified, and to obtain a measure of learners' perceived CS in specific relation to the self-concept described in the opening statement.

Confidence: This third item measured participants' 'confidence' in their current abilities. Linguistic self-confidence has been identified as an important component of language learning (Sampasivam & Clement, 2014). Similar to measuring the strength of an IS via vividness and desirability, this was designed to capture the perceived strength of a learner's current self-concept in each of the domains of Arabic identified.

Each sub-item was measured via a 6-point bi-polar Likert-scale, in which both the negative and positive ends are verbalised (as illustrated in Figure 3.5). A 6-point scale was chosen so as to encourage students to avoid simply ticking the middle column if

responses would remain anonymous. These answers were also the reason for identifying the urgency surrounding the issue of motivation amongst GCSE Arabic learners. Students in the current study were therefore given ample opportunity to express their experiences, attitudes, and feelings towards the language and their learning experience, and how they perceive their own future selves in relation to Arabic.

The incorporation of qualitative design elements into the measurement of motivation is also vital in facilitating greater analysis of learners' self-concepts and their internal mechanisms, and to ensure that important contributors and components of learners' motivational profiles are not overlooked.

Care was taken to separate questions concerning learners' attitudes towards the language itself, and the experience of taking the GCSE since findings in previous studies (Ramezanzadeh, 2015) indicated that students had varying views on the language depending on their exposure to it in different learning environments (e.g., at the mosque as opposed to studying it in school).

3.4.1.1.4. *Self-Rated Proficiency*

Students were also asked to rate their own proficiencies across an array of Arabic-related language skills on a 6-point Likert-scale to provide a comparative measure to the proficiency scores obtained via the proficiency tests (item 21). These items include the four skills of reading, writing, speaking and listening, but adopt a finer-grained approach by incorporating questions on variety-dependent language skills, namely: reading with vowels; reading and understanding religious texts; and speaking and understanding spoken colloquial varieties (Table 3.4).

Table 3.4. Language Variety-Related Skills included in Self-Rated Proficiency Scale	
Arabic Variety	Skills
Modern Standard	Writing; Speaking; Reading aloud; Reading comprehension; Listening comprehension
Classical	Reading aloud (reading with harakat/diacritics); Reading comprehension; Listening comprehension
Dialect	Speaking; Listening comprehension

The inclusion of the Self-Rated (SR) Proficiency measures and proficiency tests was designed to provide a detailed picture of learners' actual and perceived proficiency levels so as to be able to compare these with different aspects of their motivation; and to avoid the pitfalls of relying only on self-rated proficiency scores without more objective measures of learners' levels, measurements of which are currently lacking in the literature.

3.4.1.1.5. *Instructions*

The Motivation Survey booklet was completed by participants by hand. It included an introduction and instruction page informing students that they were free to skip any questions they did not feel comfortable answering, and that they could ask for help from the researcher/supervising teacher at any point during the survey. The anonymity and confidentiality of their answers was stressed in the instructions read out before the survey commenced. Since participants were taking these surveys and proficiency tests during their examination years, they were assured that none of their answers would affect their GCSE grades in any way. Learners were also asked to complete the survey quietly and independently to avoid collaboration on answers, which could then not be used in the dataset. Students were thanked before and after the survey session for their participation.

The booklets were entitled 'Arabic Language Survey' rather than 'Motivation Surveys' so as not to alert the students to the construct being measured, or pressure them to answer more positively on these questions.

3.4.1.2. Proficiency Tests

The limited existing proficiency tests for Arabic vary in their quality, scope, level, and accessibility. Statistics on the reliability and validity of these tests are sporadic, and details of their development are often difficult to ascertain. A number are only available for a fee from the online Foreign Language Assessment Directory. One study which details the development and analysis of a 'curriculum independent' Arabic proficiency test is Ricks (2015). For his doctoral thesis, Ricks developed three different lexical knowledge assessments for Arabic and examined their reliability and validity: a Yes/ No (Y/N) Test; a Vocabulary Levels Test (VLT); and a Word Associates Format/Test (WAF).

All three tests 'performed well according to traditional measures of test reliability and item performance' (p.165), and though these focused solely on vocabulary knowledge, measurements of learners' vocabulary size have been shown to correlate consistently with other language proficiency measures and thus, Ricks argues, serve as a reliable predictor of functional language ability.

Based on Rick's work, a Yes/No Vocabulary Item Test was developed for this study (see Appendix B.3). This test was chosen from the three, given that: a) some GCSE learners may have spent only a year learning Arabic and as a result, the VLT and WAF would be too difficult as they required a broader level of lexical knowledge in order to provide synonyms and to understand word definitions in the target language; and b) measurements of vocabulary size have been shown to correlate with vocabulary breadth relieving for the need for tests of both, which would likely have caused fatigue effects given the similar layouts. The isolated nature of words in these three tests did little to replicate the format of any of the GCSE questions, thus a cloze test was also developed to provide a parallel measure of participants' proficiency levels.

3.4.1.2.1. *Y/N Test*

Yes/No vocabulary tests are so called because they require participants to indicate from a list of items which words they do or do not know by entry of a 'yes/no' or a check mark.¹⁰ The test was brought to prominence by Paul Meara (Meara & Buxton, 1987; Meara, 2010) and has developed into a popular placement test. Whilst it does not provide a detailed or robust indication of the depth of a learner's vocabulary knowledge, it is considered to be a flexible and straightforward measure of a person's vocabulary size (Ricks, 2015) and one whose scores correlate well with other proficiency measures (Alderson, 2006).

3.4.1.2.1.1. *Real Words*

In order to afford ecological validity to this test format, vocabulary items were drawn from the list provided for the GCSE Arabic exam on Pearson Edexcel website. These were then cross-referenced with Buckwalter and Parkinson's Frequency Dictionary of Arabic (2011) and given a frequency ranking. Vocabulary items were then organised into rank order and separated into levels/bands from 1 to 5 (1-1000 (Band 1), 1001-2000 (Band 2), 2001-3000 (Band 3) etc.). Items in the GCSE vocabulary list which did not appear in the most frequent 5,000 words were excluded.¹¹ 60 items were chosen in total, listed in Figure 3.6.

¹⁰ A modification of this test is known as the 'checklist test', whereby participants simply check the items they do know, and leave the others blank. Despite the slight difference between the two tests, Ricks (2015) opted to use 'Yes/No Test' to refer to his test design given the term's ubiquity and the fact that the two terms are used interchangeably in the literature; this labelling was adopted in this study and the test is henceforth referred to as a 'Yes/No Test'.

¹¹ The word frequency band analysis unearthed some worrying trends in terms of the accessibility of these lexical items for learners. For a breakdown of word frequency statistics for the GCSE core vocabulary items, see the Appendix A.1.

Level 0		Level 1		Level 2		Level 3		Level 4		Level 5		Pseudowords		
Item	Word	Item	Word	Item	Word	Item	Word	Item	Word	Item	Word			
1e	لكن	1a	شارك	1c	كافي	1h	فلوس	2b	مسك	2p	كمال	يحشج	افتماء	مناوغة
2f	حتى	1b	نقل	1i	لقب	2o	صنع	2d	عبور	3f	بصل	ارتماپ	مستذر	متجامت
2n	أو	1f	الآن	1l	تجاه	2q	متأخر	2e	ضائع	3i	قطن	إشفام	لكوذ	عالر
3c	بعض	1j	موسم	1m	حاد	3a	مماثل	2g	عريس	3j	لعبة	زماء	منشرش	تبامش
3k	واحد	2l	خبر	1n	ثقیل	3d	تمثيل	2h	قلعة	3m	غضب	قماذ	ملنشة	رفخ (2x)
3p	عند	4a	علاقة	1p	متر	3e	كسب	2i	بنزين	4e	متساوي	تخميف	غمكان	تخفیب
3q	ليس	4b	منذ	4i	رمز	4h	ضاحية	4d	عائب	4n	مصروف	احتذاك	ندقة	ملفاة
4g	إلى	4j	بدون	4p	حاضر	4o	مثير	4m	حرق	4q	طبق	شلیق	مجتبئ	منطولة
5b	كبير	5c	اجتماعي	5l	مستقل	5g	ستون	5d	مرضي	5m	سكين			
5n	نفس	5p	أداء	5o	سائق	5h	ذكي	5j	مزارع					
		5q	خرج											

Figure 3.6. Y/N Test Items organised by Word Frequency Level

10 items from frequency bands 1-5 were selected for inclusion in the instrument. 10 words were also selected from the first 100 words. These were included to counteract the potential frustration felt by participants upon encountering a large number of words they may not know, given that some individuals may have only been learning Arabic for less than a year by the time the first phase was conducted. All words were selected at random and were then ordered randomly throughout the test.

Selected items were confined to verbs in their perfect third person masculine singular form (the simplest variety of the verb and the stem from which other varieties are conjugated), singular nouns, and singular adjectives. Arabic contains multiple verbal derivational patterns which can affect the appearance of the verb significantly, thus no conjugated verbs were included. Focusing on these categories avoided participants' lexical knowledge scores being confounded by their knowledge of verbal morphology. Items were left unvoiced for ecological validity since this is how they appear in the GCSE course material.

3.4.1.2.1.2. *Pseudowords*

Pseudowords are items ‘that fulfil the phonological constraints of the language but do not bear meaning’ (Huibregtse, Admiraal, & Meara, 2002, p.227). These mimic acceptable word lengths, syllable patterns and phoneme combinations in the target language but which have no meaning.

Pseudowords were originally included in the list of items to act as a safeguard against participants simply ticking all the words if they felt confident in their knowledge of the language. According to Ricks, ‘the ability of respondents to discriminate between real and false words is perhaps the best piece of affirmative evidence that they do in fact know at least something about the words they claim to know’ (2015, p.64).

The pseudowords in the Y/N test were taken from the list of plausible pseudowords developed by Ricks in his thesis (p.246-248) who organised these items into difficulty bands. Five words were randomly selected from each band, making a total of 25 words which were then interspersed throughout the real word items. Analyses were run to determine the reliability of this instrument, which was found to have good levels of internal consistency (see Appendix B.4).

3.4.1.2.2. *Instructions*

Participants were asked to read an instruction page before commencing the test. They were given the following set of instructions regarding the Y/N test: ‘If you know the word, tick the box next to it. If you don’t know the word, or are not sure, leave the box blank.’ They were then provided with an example to illustrate this. Participants were not told that the test included pseudowords.

Proficiency tests were entitled ‘Activity Booklet’. This was to lessen any perception amongst participants that they were being examined on their skills, particularly given the high-pressured GCSE environment they were studying in. These booklets were also completed by hand.

3.4.1.2.3. *Scoring*¹²

Participants were awarded one point for a 'hit' (a tick against a correct word), none for a 'miss' (no tick). Any checks made against pseudowords were ignored. Pseudowords served to alert the researcher as to whether any students had just ticked the entire list without paying attention to the words, in which case the data could be excluded.

3.4.1.2.4. *Cloze Test*

To try to ensure a more robust measure of participants' proficiency, a cloze test was developed alongside the Y/N vocabulary test (see Appendix B.3). C-tests have been the predominant go-to measure for language proficiency (Eckes & Grotjahn, 2006). These involve the systematic deletion of the end half of every nth word within a text.

This test combines lexical and morphological components (Alderson, 2002), and as such the construction of these prove difficult in the context of Arabic, since the language involves complex derivational and conjugational patterns involving multiple prefixes, suffixes and infixes, making it difficult to ascertain whether a person is unfamiliar with a stem or a conjugational/ derivational pattern. In addition, the majority of letters in their beginning and middle forms are written in connection with the following letters. Current word processing programs (still) do not allow for letters to be written in their joined form if the following letter is absent, making the removal of half a word practically impossible.

¹² There are two approaches to scoring a Y/N test. In the first, participants are awarded a mark if they indicate that they know a real word and deducted a point if they check a pseudoword. This follows the protocol designed to test the Signal Detection Theory (Macmillan, 2002). In this case, the 'signals' are the real words. If a participant indicates that they know a real word i.e., they have detected that the signal is present, and they score a 'hit'. If they fail to identify a real word, the signal is undetected, and they 'miss'. If they indicate that they know a pseudoword this is considered a 'false alarm', and if they do not check a pseudoword this is considered a 'correct rejection'. The second is the method outlined above, with points awarded for a hit, and marks against pseudowords ignored. Given the already low levels of vocabulary knowledge amongst some learners, it was decided that deducting points for 'false alarms' may in fact result in some participants scoring very low, possibly negative scores despite having some knowledge of real items. Thus, the second approach to scoring the Y/N test was employed.

To retain the ecological validity of using text-based questions, a cloze test was developed as an alternative. This involves the full deletion of every n^{th} word in an appropriate text or series of texts. For the purposes of this study, texts were selected from previous GCSE exam papers.

Cloze tests have been shown in the past to be a reliable measure of learners' proficiency levels in other languages (Stubbs & Tucker, 1974). Since no appropriate existing cloze test of Arabic was readily available, one was designed for this study. The test was developed as per recommendations outlined by Toeimah (1978) in constructing his own proficiency tests for Arabic for university-level learners in the US.

3.4.1.2.4.1. Text Selection

Texts were carefully selected from pre-existing reading texts from GCSE past-papers to ensure they were representative of students' usual material. A range of passages were chosen in order to include a variety of subjects. Low-mid grade passages were selected to avoid floor effects, but to ensure a reasonable range of difficulty, since students in Year 10 may have studied the language course for less than a year. Texts were selected from the 2017 sample papers for Unit 3: Reading and Understanding (Pearson, 2017), which students were unlikely to have had previous experience with, given it had only recently been released at the time of testing. Three texts were selected, two from the Foundation Tier and one from the Higher Tier paper so as to provide a range of difficulty levels. To try to eliminate the chance of floor effects, diacritical marks were also added by the researcher to the first text to facilitate participants' reading of the script. The number of items in each text is outlined in Table 3.5.

Table 3.5. Cloze Test Text Sources			
Text	Source	Content	No. of Items
1	2017 Sample Paper Foundation Tier, Question 1	Daily Routine (vowelled text)	8
2	2017 Sample Paper Foundation Tier, Question 10	Leisure Activities	5
3	2017 Sample Paper Higher Tier Question 3	Travel	12

3.4.1.2.4.2. Constructing Test Format

Every seventh word was deleted, as opposed to every five words, in order to make the test more accessible to non-native speakers (Klare, Sinakiao & Stulran, 1972). The structural deletion method was applied, in which words are deleted regardless of their grammatical function, since there is evidence to suggest that this method affords better discrimination than lexical deletion (Taylor, 1956). 3 multiple choice items were included for each answer. Free-response items have most commonly been employed in cloze tests, however Cranney's study comparing free-response and multiple choice, concluded that the latter constituted a better measure of proficiency (1968). Although this method fails to test students' production, it reduces discrepancies and complications involved in the scoring system, in which the acceptability of written answers depends ultimately on the marker's judgement. Jonz (1976) also claims that a multiple-choice cloze test aids the student in completing the task. Alderson (1990) also reported that providing the multiple-choice method lessens the memory load of the test-taker and making completion easier.

3.4.1.2.4.3. Selection of Multiple-Choice Distractors

Two main methods of selecting distractors have been used in research conducted to date: the empirical method, and the rational selection method. The former uses common errors derived via pretesting of a completion cloze test. The rational selection method involves experts selecting distractors based on a consideration of possible sources of examinee error from a reading of the passage (Hale et al., 1988).

The rational selection method has been shown to be at least as effective in measuring proficiency as empirical selection (Scholz & Scholz, 1981). Thus, this method was implemented and the distractors generated by the researcher based on previous experience taking, teaching and researching the Arabic GCSE.

It was decided, based on an estimate of students' levels, having only studied a term of GCSE level Arabic, that distractors would be lexically varied only, and not grammatically varied. Thus, for example, distractors for a deleted definite, singular, feminine noun

consist of only definite, singular, feminine nouns. That is not to say that this only tests learners' lexical knowledge. According to Hale et al. (1989), 'It is possible that the kind of text processing required by cloze items in general may be more holistic in character than current linguistic analyses might suggest. In order for the reader to deal with a question involving grammar, for example, vocabulary must play some role; and one cannot answer a question involving vocabulary without a certain control of grammar' (p.65). Reliability analyses demonstrated the internal consistency this instrument (see Appendix B.4).

The final printed tests were entitled 'Activity Booklet', rather than 'Proficiency Test' so as not to unduly stress students giving them the impression that they were being formally tested or examined.

3.4.1.2.4.4. Instructions

Participants were given the following set of instructions in regard to the cloze test: 'On the following pages are some sentences in Arabic. Your job is to pick the underlined word that fits each sentence. Please circle your choice.' They were then provided with an example to illustrate how to indicate their chosen answers.

3.4.1.2.4.5. Scoring

Participants were awarded one mark for identifying the correct word, and no points if they identified the wrong word; circled more than one word; or failed to indicate an answer at all.

3.4.1.3. Interview Protocols

The interviews followed a semi-structured protocol, allowing for flexibility in approach and direction (Savin-Baden & Howell Major, 2013). The prompts (outlined in the Interview Protocol in Appendix B.5) were designed to elicit information concerning participants' Arabic learning experiences, and their motivational self-concepts:

Learning Experience

Learners were first asked to describe their experiences learning Arabic and studying the GCSE so far. This was included as the first question to encourage students to reflect on their learning, and to help them feel comfortable with the interview setup by discussing their general views and feelings towards the subject.

Current Self

Learners were then asked how confident and/ or comfortable they felt with the language at present and whether or not this had changed over the year. This was designed to ascertain information about their CS concepts in regard to Arabic.

Ideal Self

General

Following a discussion of their current levels of confidence and their experiences with the language, learners were asked to imagine themselves 5-10 years in the future, and to describe their general future self-image. The purpose of this question was to encourage learners to switch their attention to the future, rather than the present, before asking them about their Arabic-related self-images. This also allowed the researcher to identify whether learners were capable of imagining their future selves, since some research suggests that younger learners are not as adept at engaging in this vision-generating, elaboration practice (Dörnyei, 2009; Shapka & Keating, 2005; Zentner & Renaud, 2007).

Arabic-specific

Learners were then asked whether or not Arabic played a role in their future self-images, and if so, how it was incorporated. Where learners did describe an Arabic IS image, follow up questions concerning frequency, vividness, and plausibility were asked. They were also asked whether their IS images had changed over the course of the year.

Ought-to Self

Participants were next asked about their OTS, specifically whether or not they felt 'pressured' in any way to learn Arabic. This wording was considered to be more transparent than asking learners whether they felt they 'ought' to learn the language. Prompts included questions regarding where this pressure came from, whether there were any consequences involved in not achieving an Arabic-related future self-image, and whether or not this pressure had changed over the year.

Religion

A separate question on potential religious motivations was included to examine how this component fits into learners' self-concepts, since this has not been explored in previous L2MSS studies but is pertinent to the study of the Arabic language. Learners were asked whether or not religion played a role in their learning of Arabic, and where it did, to elaborate on this role. This was not explicitly linked to their questions on the CS, IS or OTS since this is a novel addition in the investigation of the L2MSS, and the researcher was keen to explore all possible dimensions of this factor in learners' motivations towards Arabic.

Learners were also given the opportunity at the end of the interview to add any additional comments they wanted to make regarding Arabic and/ or the GCSE and to ask any questions.

3.4.2. Piloting

The Motivation Survey and Proficiency Tests were cognitively pre-tested on three separate occasions. Cognitive pre-testing involves the 'administration of draft survey questions while collecting additional verbal information about the survey responses, which is used to evaluate the quality of the response or to help determine whether the question is generating the information that its author intends' (Beatty & Willis, 2007, p.287).

3.4.2.1. Pilot 1

The first pilot was conducted with a class of 16 students studying Arabic in Year 10 at a girls' Islamic school. Participants were asked to complete the survey and proficiency test, and then provide verbal feedback on the comprehension and ease of answering the questions. The following modifications were made:

The question on self-reported predicted grades was removed since, being at the beginning of Year 10 when the first survey was developed, most students did not know their predicted grades, and the production of ad hoc estimates by the teacher seemed uninformative.

The item concerning languages spoken in the home was also modified. Originally, this was presented as a list of check boxes with the options of English and some common heritage languages in the participant community. Given the diversity of responses to the 'Other' option and the intricacies of language usage, expressed by responses such as: *'English, and Urdu in my grandparents' house'* it was decided that this would be better worded as an open-ended item. This would avoid presenting some languages as more mainstream than others, and encourage students to consider the dominant languages used in their homes, commenting on the different types of usage where necessary.

The wording of the motivation scale items was considered to be clear for all items except the OTS statements referring to obligation/ force. The original wording 'I feel like/ don't feel like I have to achieve this' was not closely related enough to the concept which it was measuring. The wording was altered to 'I feel forced/ don't feel forced to achieve this', thus clarifying the sense of obligation being measured in the item.

Originally, an 11-point Likert-scale had been adopted to match Hessel's (2015). In view of the number of students who opted for the middle point on the 11-point Likert-scale or who opted for ticking the same box for all items, it was decided that this number of scale points was clearly not conducive to eliciting accurate responses. A shorter, even-numbered scale, in this case a 6-point scale, was adopted to both avoid the potential 'I don't know/ care' option of the middle point, and to make the items less overwhelming

in terms of choice. The tick circles were also graduated in size and printed in colour to make the scales more appealing to the eye.

The original version of the proficiency tests consisted of a cloze test with 42 multiple choice items, based on texts taken from the higher ends of the GCSE Reading and Comprehension past papers. It was clear from the feedback from the students after completing these, that these were too difficult and that this section of the test was too long. As a result, other test formats were explored; the Y/N test was developed, and a revised, shortened cloze test formulated for further piloting.

It also became clear from the participants' feedback that the overall survey and test materials themselves needed to be engaging and appealing. The questions were subsequently reformatted into booklet form, and printed with a brightly coloured, purposely designed cover to make them feel less like 'forms' or the standard exercises or tests learners may receive in class.

3.4.2.2. Pilot 2

Following modifications from the first pilot, the surveys and proficiency tests were cognitively pre-tested with a group of female students aged between 11-18 years during a summer camp at which the researcher was mentoring. Students were asked to complete the instruments individually, and ask any questions they had whilst filling out the booklets. Students were then asked at the end of the session for any feedback regarding the ease of completion of the booklets and their comprehension of the questions. It was clear from this pilot phase that the reformatting of the booklets had been successful in making them more appealing to younger learners, since they had no trouble completing the booklets.

The questions on self-rated proficiency were moved from being the first set of scale questions to later in the list of items. This was intended to make the survey less intimidating, allowing participants to reflect on their future self-images without first undergoing an appraisal of their proficiency levels, which it was felt may have a demoralizing effect where students did not feel competent in their current skills. Had

the motivation survey not been followed by the proficiency test, it would have been most effective to place this set of scale items at the end of the booklet. However, given the responses of students in Pilot 1, who seemed very discouraged when it came to their perceived current abilities, it was felt that this may affect learners' levels of enthusiasm to complete the Activity Booklet. As such, it was placed just past the half-way mark of the survey.

The proficiency tests themselves were also completed in a straightforward manner without the fatigue effects obvious amongst participants in the previous pilot phase. The results exhibited a range of proficiency amongst participants, who came from a wide range of levels and backgrounds (both ethnic and educational). Whilst it was not possible to run any inferential statistical analyses on these results, based on a visual examination of the data, fears of floor or ceiling effects were allayed.

3.4.2.3. Pilot 3

Following these alterations, both booklets were finally piloted with a class of first year university learners of Arabic to test for clarity, comprehensibility, and any overlooked issues with item phrasing. Participants were asked to feedback verbally after completing each booklet, and to make notes on the booklets themselves to record any issues. A number of points regarding the wording of questions related to demographic information were raised, and the booklets modified accordingly before being printed.

Given time and logistical constraints, it was not possible to pilot these with a group of learners from the same sample large enough to empirically validate the instrument. Thus, the empirical validation of these instruments constitutes a primary research question of this study.

3.4.3. Recruitment

Following the initial composition of the schools database, all schools were contacted by first by phone, and then email with a written overview of the study, and invited to participate in the research. In schools that expressed interest, agreement was obtained

from the school's Head Teacher or Head of Department, and the GCSE Arabic teacher. A total of 16 schools responded positively and agreed to participate in the first year of the research. In three of these schools, Arabic was taught across two Year 10 classes. In one school, students were divided arbitrarily into two classes, taught by different teachers. In two other schools, students were divided by gender, also taught by different teachers. This meant a total of 19 different cohorts of learners were recruited to participate in the study. An overview of these schools is provided in Table 3.6.

Students and their parents were provided with consent forms and informed of their right to decline participation in the study. In this instance, opt-out consent forms were provided since only written survey response items and proficiency test data was being collected. Teachers were tasked with collecting any return slips. Students who returned either form were not included in the data collection.

3.4.3.1. Participants

3.4.3.1.1. *Schools*

Participants were recruited from the population identified in the initial database compilation, which consists of 144 schools. The majority of these are Muslim faith schools (77.8%), with the remaining 22.2% a mixture of other faith and non-denominational schools. A total of 19 Arabic classes drawn from 16 different school establishments agreed to participate at T1. The schools were situated in 9 different cities in England, with the majority (9 schools) situated in London.

Four of these schools were non-faith schools, however students' responses to items concerning religious affiliation indicated that the vast majority (95.9%) were from Muslim backgrounds, despite the diversity in school types. These figures are explained by the fact that, even in non-faith schools, pupils from a Muslim background are likely to be attracted, or encouraged, to take up the study of the language given its ties to their faith.

Of the participating schools, two were officially registered as 'Mixed' according to their most recent Ofsted reports. However, in one of these cases (Classes 4 and 5), male and female students were taught separately, in two different classes by separate teachers. In another (10 and 11), the year 10 cohort was divided into two cohorts arbitrarily, again taught by two separate teachers. Only one school (14) was co-ed in the traditional sense, with male and female students being taught in the same classroom. In the interests of clarity therefore, class cohorts rather than school cohorts will be referred to.

Table 3.6. Overview of Schools and Sample

School Variables						T1		T2		
School Code	Location	School Type	Faith	Gender	Data Collection Method	Mtv Survey	Prof Tests	Mtv Survey	Prof Tests	Interviews
1	Manchester	Independent	Faith (Muslim)	Girls	By post	11	11	10	10	
2	Luton	Independent	Faith (Muslim)	Boys	In person	9	9			
3	Luton	Independent	Faith (Muslim)	Girls	In person	9	9			
4	London	Independent	Faith (Muslim)	Mixed (Girls Class)	In person	10	10	10	10	10
5	London	Independent	Faith (Muslim)	Mixed (Boys Class)	In person	18	18	12	10	15
6	London	Independent	Faith (Muslim)	Girls	In person	15	15	14	14	3
7	Bolton	Voluntary Aided	Faith (Muslim)	Girls	In person	27	27	26	26	
8	Bradford	Academy	Non-Faith	Girls	In person	14	13	14	15	12
9	Bristol	Academy	Non-Faith	Girls	In person	3	3			
10	London	Free School	Faith (Muslim)	Girls	In person	27	27			18
11	London	Free School	Faith (Muslim)	Girls	In person					
12	Bradford	Academy	Faith (Muslim)	Girls	In person	18	18	12	11	10
13	Hertfordshire	Independent	Non-Faith	Boys	In person	7	7	7	7	6
14	Sheffield	Community	Non-Faith	Mixed	By post	13	12	12	12	
15	London	Independent	Faith (Muslim)	Boys	In person	25	25	25	25	
16	London	Independent	Faith (Muslim)	Girls	In person	7	7	6	6	6
17	London	Independent	Faith (Muslim)	Girls	In person	21	21	19	19	15
18	Rochdale	Independent	Faith (Muslim)	Girls	By post	7	7			
19	London	Independent	Faith (Muslim)	Boys	By post	24	24	8	8	
TOTALS:						265	263	175	173	95

At T2, four schools ceased their participation, either citing school timetable pressures or simply ignoring correspondence. Three of these schools were girls' schools, one of which included two classes, and one was a co-ed school in which the classes were gender segregated. The number of classes participating thus fell to 13. Only one of the schools to withdraw at T2 was a non-faith school, the rest were Muslim schools. In several schools, a number of participants were also either absent or declined to participate. Overall, therefore, the number of participants at T2 was 178.¹³

3.4.3.1.2. *Students*

Student participants were all in Year 10, aged 15-16 years, and were being taught Arabic for a minimum of an hour a week. 265 Year 10 Arabic students participated at T1, and 175 at T2 (see Table 3.6), with 95 taking part in interviews.

3.4.3.1.2.1. *Sample Descriptives*

Gender

At T1, 65.3% of student participants were female, and 34.7% were male. This changed by only 1 percentage point at T2 (see Table 3.7).

Table 3.7. Gender Distribution of Sample at T1 and T2 Surveys								
	T1				T2			
	Classes	Students	Total	%	Classes	Students	Total	%
Girls	13	169	173	65.3	8	112	116	66.3
Boys	5	83	92	34.7	4	51	59	33.7
Mixed	1	13 (G=4; B=9)			1	12 (G=4; B=8)		

Religion

The number of participants of religions other than Islam was 3 at both timepoints. At T1, one identified themselves as Christian, one as Buddhist, and one as Other. At T2, the

¹³ The total number of students was 178 at T2, of these 175 took the surveys and 173 took the proficiency tests. Some participants took only the survey and some took only the proficiency tests, hence the difference in totals.

same participant identified as Christian, while the participant who had previously ticked Buddhist, ticked 'None'. The participant who had ticked 'Other' at T1, identified themselves as Muslim at T2, however a participant who had previously identified themselves as Muslim, at T2 ticked 'None'.

Heritage Status

Classes contained students from a wide range of ethnic and linguistic backgrounds. Since the vast majority of participants were from Muslim backgrounds, an ancestral definition here was deemed to be unhelpful. To distinguish heritage learners (HLs) from non-heritage learners (NHLs), a variation of Valdés' (1999) definition was employed. This included participants who fell into one or more of the following categories:

- Students who identified their ethnic origin as 'Arab', or 'Mixed' and specified an Arab country of origin;
- Students with one/ both parents of Arab origin;
- Students who had lived in (not just visited) an Arabic-speaking country;
- Students who identified Arabic as being spoken in their home.

On this basis, 65 students (24.5%) were classified as HLs at T1, with the remaining 200 constituting NHLs. Of the 65, 41.5% were of Arab ethnicity. Of the remaining 38, 12 were of mixed ethnic origin, and 15 were of Black African ethnicity.

The number of HLs at T2 was 48 out of a total 175, however the proportion (27.4%) remained roughly the same between the two timepoints. A full explanation of the procedure for assigning participants to the heritage or non-heritage learner category is provided in Appendix B.6.

Language Background

Participants listed 25 languages other than English as spoken in their homes. It appears that the majority of participants came from multilingual backgrounds, with only 19 out of 265 (7.2%) indicating English as their only home language. 213 participants (80.4%)

indicated that languages other than English was spoken in their homes, either in isolation or alongside English, and 47 of the 265 (17.7%) listed two or more languages (other than English) in their response at T1. Of the languages, Arabic was the most commonly mentioned (18.5%), followed by Bengali (17.7%), Urdu (16.6%), and Somali (15.8%). The distribution of major home languages stayed relatively stable between the two timepoints. Table 3.8 lists the major languages identified at T1 and T2. Other languages listed included Farsi, Hindi, Pushto, Hindko, Swahili, Italian, French and Dutch.

Table 3.8. Home Languages of Participants				
	T1		T2	
Language (Other than English)	Freq.	%	Freq.	%
Arabic	49	18.5	30	17.1
Urdu	44	16.6	30	17.1
Punjabi	13	4.9	16	4.9
Gujurati	28	10.6	15	9.1
Somali	42	15.8	22	12.6
Bengali	47	17.7	20	11.4

Previous Arabic Study

The vast majority of students (96.7%) had studied Arabic for at least a year prior to commencing their official KS4 study in Year 10. Almost half of them had been studying it for three years, and 77% had been studying it in school for up to four years. This is explained by the fact that most Muslim secondary schools offer Arabic from Year 7, and those students who also attended a Muslim primary school will have begun their study even earlier. This pattern is repeated at T2, with a very small minority of students having started learning Arabic in Year 10 (4.1%), over 50% of participants having studied Arabic for at least 3 years (see Table 3.9).

Table 3.9. Number of Years of Arabic Study prior to Year 10						
	T1			T2		
Years	Freq.	%	Cum. Freq. %	Freq.	%	Cum. Freq. %
<1	9	3.4	3.4	7	4.1	4.1
1	11	4.2	7.6	10	5.8	9.9
2	16	6.1	13.7	14	8.1	18.0
3	128	48.5	62.2	60	34.9	52.9
4	39	14.8	77.0	45	26.2	79.1
5 or more	61	23.0	100	36	20.9	100

3.4.4. Data Collection Procedure

3.4.4.1. Motivation Surveys and Proficiency Tests

Following receipt of agreement from each school, information letters and consent forms (Appendix B.7) were emailed to the Arabic teacher to be disseminated to Year 10 Arabic students and their parents. The teacher was asked to collect any returned opt-out slips and inform the researcher of how many students would be participating before data collection commenced. Alternative provision was made by the teacher for any students not participating.

The researcher then liaised with the respective Arabic teachers to set a date to visit and carry out the first phase (T1) of this research. In cases where schools could not set a fixed date with the researcher in advance, the option was provided for them to complete the survey under teacher supervision. In the latter case, teachers were provided with an instruction protocol and check list to follow and complete (see Appendix B.8). Materials were posted to these school, and the teachers were asked to return booklets by post, along with the check list with confirmation that the instructions were followed and with notes of any concerns raised during the process.

Where the researcher was present, the session began with an introduction and disclaimer that the data would be anonymised. Learners were asked to put their names on the front of the booklets for administrative purposes, since booklets from each timepoint needed to be matched. An opportunity to ask questions was provided before

the surveys were distributed. Participants were asked to complete the questions independently and quietly so as not to disturb their peers. They were requested to read the instructions at the beginning of the booklet and then commence simultaneously. Participants were able to raise their hands and ask questions of the researcher or supervising teacher throughout. Participants were asked to indicate to the researcher or teacher once they had completed the motivation survey, which was then collected. They were provided with a proficiency test which they completed under the same conditions. Those participants who completed both booklets were asked to remain quiet whilst completing an activity of their choosing (e.g., homework, personal reading, 'snoozing').

Once all participants had completed both booklets, the class was given the opportunity to talk to the researcher and ask questions more generally about the study itself, the Arabic language, university life, and the researcher's own personal experiences. Teachers were often very keen for participants to use this opportunity to discuss the importance of languages and the experience of being a Muslim woman at university. Following the completion of this first part of the research, the participants and Arabic teachers were thanked and reminded that this phase would be repeated later in the year.

For the repetition of this phase (T2), each participating school was contacted individually in the final term of the school year, with an invitation to engage again in the data collection, with the option to have either the researcher present, or to complete the booklets under teacher supervision. The procedure for T2 followed the same order and instructions as T1, with the addition of an optional interview stage, outlined below.

Given timetable restraints, and other scheduling or administrative issues, four schools did not complete this phase of the research, leaving a total of 12 schools, and 13 classes participating.

3.4.4.2. Interviews

In addition to the surveys and proficiency tests, participants were also invited at T2 to participate in interviews with the researcher. 9 of the 13 classes at T2 schools agreed to participate in this component of the research.¹⁴ Since these were audio recorded and participants were under 16, opt-in consent forms were provided to parents and students before the researcher's visit. Those learners who received parental consent and were willing themselves to participate in the interviews were included in this stage of the research. In total, 95 students agreed to participate.

The interview sessions were conducted after the two booklets had been completed, during the same school visit. Interviews were conducted in small groups of 2-6 students. This approach was adopted both for expediency, given the short timeframe which the researcher was allocated to complete these during the school timetable (usually between 1-3 hours), and in order to help the students feel at ease during the interview. This approach was adopted, in line with other studies with adolescents. It was felt that students would be more comfortable participating in interviews alongside their peers, rather than one-on-one with the researcher, which may have made them feel intimidated and/or more willing or pressured to give answers they felt were 'appropriate' or sought-after in the eyes of the researcher, thus combating, to some degree, the 'performance effect' (Lamb, 2009). Whilst the existence of their peers is not guaranteed to alleviate these potential feelings entirely, conducting group, rather than individual interviews was considered to aid in some way to addressing the power dynamics between the interviewer and interviewees, reducing levels of inhibition (Field, 2000) and encourage freer discussion among participants (Lanvers, 2016).

The interview was commenced with a brief introduction by the interviewer. Students' right to withdraw themselves from the study at any point without any negative consequences was highlighted, and an emphasis was placed on the anonymity and confidentiality of all data collected. Participants were reassured both in writing and verbally that opinions expressed in the interviews would not affect their GCSE or class

¹⁴ One school which agreed to participate in the interview phase also agreed to participate in the T2 survey/proficiency test components, however, unfortunately due to a mix-up on the part of the researcher, the booklets were not completed at this school.

grades in any way. The interviews were audio recorded on a digital recording device, visible to the students at all times. Students were asked to state their names at the beginning of the interview so that the researcher could later match their responses to the survey data.

3.4.5. Data Transformation and Treatment

3.4.5.1. Motivation Surveys and Proficiency Tests

Data from the pen-and-paper surveys and proficiency tests were transferred into Excel spreadsheets. Digitisation was completed with the help of a data entry assistant (DEA), funded by a grant from the Qatar Foundation International. Data was subsequently cross-checked by the researcher to ensure for accuracy and completion. Once completed, data was imported to SPSS (version 25) and R (version 1.2.5033) for relevant analyses, outlined in Table 3.10.

Likert-scale entries which did not conform to the 'one box' rule, either because participants put ticks between two boxes, or ticks in two boxes, were removed from the dataset and treated as missing.

In order to preserve participants' anonymity, all names and identifying information were removed and each participant was given a unique ID formulated from the initials of the school and a number based on the order in which the data was digitized.

3.4.5.2. Interviews

Data collected from the interviews was transcribed into Word, using the Audacity program to play back recordings, by a DEA. Transcripts were cross checked against the recordings by the researcher. The transcripts were then entered into NVivo, along with the open-ended survey item responses.

3.4.6. Data Analysis

3.4.6.1. Quantitative data

Once the data was transferred to SPSS, it was analysed for missing data which was then imputed, using a Multiple Imputation method. Descriptive statistics were generated to examine for normality and outliers were examined in order to determine the appropriate statistical testing methods. A full overview of the methods used is included in Table 3.10.

To assess the structural validity of the motivation and self-rated proficiency survey scales, confirmatory factor analyses (CFA) were run. Cronbach's Alpha scores were generated to explore the reliability of the items involved. Descriptive statistics were also generated for the proficiency test data. The measurement quality of these instruments was then assessed by calculating Cronbach's Alpha scores per item.

To determine the nature of learners' motivational profiles, average-scores were generated for each construct and compared using independent-samples T-tests. The dataset was divided into sub-groups according to gender, and HL status and their respective average-score means compared. Paired samples t-tests were conducted in order to determine the nature of any longitudinal changes.

In answer to the question pertaining to the relationship between motivational self-concepts and proficiency both within and between timepoints, the factor scores from the CFA were used to run path model analyses. This was repeated for the scores for each variety, as well as these scores combined. Emerging classroom-level differences were examined by calculating intra-class correlation coefficients and examining caterpillar plots. The role of self-discrepancy was also explored using path modelling in order to investigate the relationship between different variations of self-discrepancy and possible outcome variables.

Table 3.10. Summary of Data Analysis Methods

Research Question		Analyses	Method	Software
1	What is the structural validity of: - the L2MSS construct as it relates to Arabic, as measured by the motivation survey scales? - the multiglossic self-rated proficiency construct, as measured by the self-rated proficiency scales?	Missing data analysis Normality and outliers Structural validity analysis Reliability analysis	Multiple Imputation method Generation and investigation of descriptive statistics Confirmatory Factor Analysis Cronbach's Alpha	SPSS R (mice) SPSS R (lavaan)
2	What is the nature of Arabic learners' motivational self-concepts? How do learners' motivational self-concepts vary based on: - Arabic language variety; - Gender; - Heritage Learner status?	Between group comparisons of construct average-scores Analysis of interview codes	Independent samples t-tests of average-scores Comparison of codes per sub-group	SPSS NVivo
3	How do learners' motivational self-concepts develop over time?	Within group comparisons of construct scores (T1 + T2)	Paired samples t-tests of average-scores	SPSS
4	What is the relationship between proficiency and self-concepts - Within timepoints? - Between timepoints?	Modelling of relationships between self-concepts and proficiency Analysis of between-classroom differences	Factor score path models Intra-class Correlation Analysis Generation of caterpillar plots	R (lavaan) R (ICC) R (lme4)
5	What role does self-discrepancy play in the L2MSS?	Modelling of relationship between self-discrepancy and outcome variables	Factor score path models	R (lavaan)

3.4.6.2. Qualitative data

A series of iterative coding cycles were carried out on the interview data and open-ended survey items. Saldaña refers to this as ‘sequential coding’ during which one coding approach is applied, followed by another (2012, p.63). Given the aims of this research, which were to investigate the validity of the L2MSS within a new and unique context, a combination of inductive and deductive approaches to the coding of the data were employed so as to provide a rigorous account of learners’ motivations and their dynamics within the Arabic GCSE context.

3.4.6.2.1. *Phase 1—Descriptive Analysis*

First, an initial phase of ‘intensive reading’ of the data (Gibbs, 2012) was carried out, during which transcripts were read multiple times, firstly along with the audio recordings, and then alone in order to increase initial familiarity with the data. No data was coded during this phase.

This was followed by a round of descriptive, holistic coding. Holistic coding ‘assigns a single code to a large unit of data in the corpus—often used as a preparatory approach—most applicable when the researcher has a general idea about what to investigate’ (Miles, Huberman & Saldaña, 2013). The data was coded according to a set of broad *a priori* codes—each of which described the general topic of a section—derived primarily from the interview questions protocol as a basic, initial coding framework (see Table 3.11). This is also referred to as the ‘start list’ method (p.83).

Data for each of these codes was then investigated as a dataset, annotated, and suggested sub-codes noted in memos. These notes and annotations were then reviewed, and additional codes generated. The data for each of these broad codes was revisited and coded again with these new subcodes.

Table 3.11. Phase 1 Descriptive Codes			
Code (and sub-codes)			Description (Interview Question/ Prompt)
Experience			What has been your experience with studying Arabic so far?
Confidence	Current		How confident do you feel with Arabic at the moment?
	Changes		Has that changes at all (over this year)?
Future Self-Concept	General		How do you see yourself 5-10 years from now?
	Arabic	Concept	Do you see yourself using Arabic in the future? In what way?/ How do you see yourself using Arabic in the future?
		Changes	Has that changed over this year?
Pressure			Do you feel any pressure to learn Arabic?/ Do you feel pressured in any way to learn Arabic?
Task Utility			Do you feel like the tasks you do are helping you with what you want to do with Arabic in the future? How?
Religion			What role does religion play in your learning of Arabic?
Motivation			General—why are you studying/ what keeps you studying Arabic/ why did you choose Arabic?
Other Factors			Classroom/ personal factors/ comments on the GCSE in general: e.g., Is there anything in your classes that affects how well you are doing/ how engaged you are?

3.4.6.2.2. Phase 2—Exploratory Analysis

Given the novel context of application for the L2MSS and the multiple issues that have been identified with the model, the data was also explored without the imposition of the above *a priori* codes. The data was revisited after a break, and each interview transcript approached without any prior codes generated. All previous codes were hidden from the data. It cannot be claimed that this is a purely inductive approach since prior knowledge of the model and first cycle codes inevitably persist and will influence, to some degree, the recoding of the data. Rather it can be said that this cycle followed ‘a theoretically driven inductive approach, in which the coding system is generated inductively, but the [researcher] draws from the theoretical perspective when identifying and naming themes’ (Syed & Nelson, 2015, p.378). In other words, where

themes from the L2MSS did emerge, these were labelled, but the researcher was free to generate her own codes unrestrained by the existing model.

The codes from the two phases were then compared for overlap, and a synthesised set of codes generated. This mirrors Saldaña's description of a 'second cycle' of coding, during which codes are reorganised into patterns or categories. This two-phase iterative process allowed for an examination of the scope and nature of the theoretical constructs proposed by the L2MSS along with a broader examination of the data with a less restrained eye so as to identify unique themes and issues arising in relation to learners' motivations to study the Arabic GCSE.

3.4.6.2.3. *Phase 3—Trajectory Analysis*

During the coding of the data for the second phase, it became clear to the researcher from the emerging themes that—given the complex nature of motivation, and the importance of examining learners' individual trajectories—the data for each participant should be examined in isolation of their peers. The data was thus re-organised by participant (rather than interview group). This represents a 'case-oriented approach' in which profiles of individual participants are compared, as opposed to the prior two phases which constituted a 'variable-oriented approach' (Miles et al., 2014). Cases were examined individually and coded for existing and any emergent themes. In particular, relationships between individual's learning experiences and their self-concepts were examined to identify whether themes at the group and school level were existent at an individual level. Codes generated in this cycle were added to the coding framework.

Codes were then collated, synthesised (whereby overlapping codes from each phase were collapsed) or expanded (by dividing a code into multiple sub-codes), in a final round of 'second cycle' coding. The full list of codes and descriptions can be found in Appendix B.9.

3.4.6.2.4. *Inter-Rater Reliability*

Codes were subjected to reliability checks to explore the consistency and replicability of the coding framework developed (Syed & Nelson, 2015). The researcher enlisted an inter-rater to code a percentage of the data according to the coding framework (Appendix B.9). The framework was first explained to the inter-rater, who then coded a percentage of the data using the codes provided. Then, inter-rater reliability was checked by comparing the inter-rater's coding with the researcher's own and calculating the percentage of agreement.

A total of 100 statements were randomly selected from the interview data. This represented approximately 10% of the coded segments from the interview transcripts. In total, the inter-rater coded the 100 statements with 164 codes. The researcher's own coding choices corresponded with 156 (95.1%) of the codes selected. In total, the researcher coded the same statements with 198 codes. The inter-rater's level of overlap of these codes was therefore 79.1%. Both these figures are deemed acceptable and above the 70% minimum agreement threshold required for inter-rater reliability (Hartmann, 1977; Stemler, 2004).

3.4.6.3. *Following the Thread*

A common charge levelled at mixed-methods studies is that they fail in delivering an integration of the strands of data collected, thus making them, at most, a form of multi-method studies, or those which involve a 'rather superficial mixing of relatively independent qualitative and quantitative components' (Boo et al., 2015, p.156). To avoid this, and to fully harness the power of a mixed-methods approach the data analysis was strategically driven by adopting a 'Following the Thread' approach (Moran-Ellis et al., 2006). This involves selecting a particular finding or 'thread' from one dataset and following it across another, constituting an 'iterative process of data interrogation which aims to interweave the findings that emerge from each dataset' to 'create a constellation of findings which can be used to generate a multi-faceted picture of the phenomenon' (p.54).

The approach requires that each dataset be analysed first independently. Next, interesting themes, loose-ends, or unanswered queries are identified, and a question generated. The other dataset is then examined for answers to this question. In some instances, this re-examination of the data also raises interesting findings or queries which are then carried back over to the first dataset, which is subsequently re-examined for further insights. As a result, a rich tapestry of findings is presented, woven together with strands from both the quantitative and qualitative components.

4. FINDINGS

The focus of this study was to examine the validity and application of the L2 Motivational Self System (L2MSS) in identifying the nature of learners' motivations towards studying Arabic—a multiglossic language—and the relationship with learners' proficiency levels in the context of UK secondary schools, where students were studying for the General Certificate of Secondary Education (GCSE). Five research questions form the structure for this study, summarised in Figure 4.1.

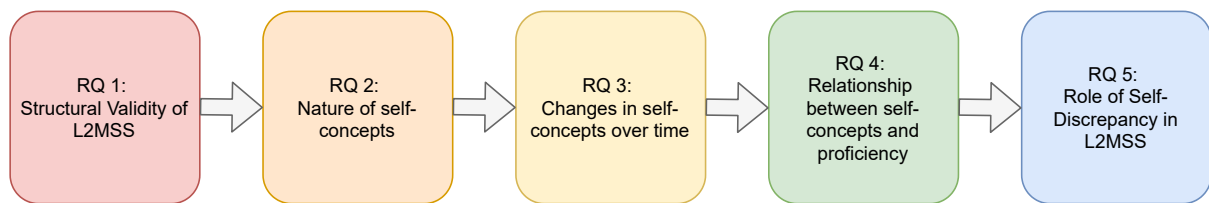
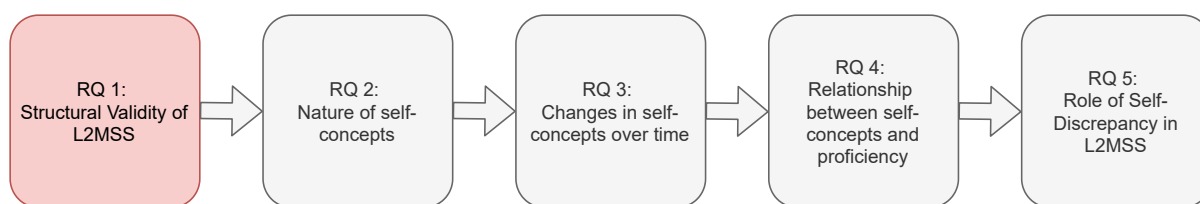


Figure 4.1. Summary of Research Questions¹⁵

The purpose of these questions was to determine, first, the structural validity of the L2MSS in the aforementioned context. This was established through confirmatory factor analyses (CFA). The second research question, for which the findings are most extensive, focuses on exploring the nature of the self-concept components of the model (Current, Ideal, and Ought-to Selves) as they emerge for each Arabic variety (Modern Standard, Classical, and Dialect) through comparison of mean scores and investigation of the interview data. Changes in learners' self-concepts and levels of proficiency between the two timepoints (at the start and end of the academic year) were explored in response to the third question. Next, through Factor Score Path Modelling, the interplay between self-concepts and three measures of proficiency (two objective: Yes/No and cloze tests, and one self-report) were investigated within and across timepoints to identify the relationship between the L2MSS components and these motivational outcome measures. The workings of this mechanism, in particular the role of self-discrepancy, are explored through the final research question.

¹⁵ This figure is repeated at the beginning of each findings section to sign-post the progression of the chapter.



4.1. Structural Validity of the L2MSS and the Self-Rated Proficiency Scales

Research Question 1:

What is the structural validity of:

- a) the Second Language Motivational Self System as it relates to Arabic, as measured by the motivation survey scales?*
- b) the multiglossic self-rated proficiency construct, as measured by the self-rated proficiency scales?*

In order to first determine the structural validity of the underlying L2MSS model used to measure learners' motivations, confirmatory factor analyses (CFA) were employed to examine whether the Likert-scale items loaded onto the hypothesised latent self-concept variables: Current, Ideal, and Ought-to Selves. This was done for each of the sets of items answered by learners in response to three opening statements, capturing the Modern Standard, Classical, and Dialect varieties of the Arabic language. The validity of an 'Overall' higher-order construct model (combining item scores for all varieties)¹⁶ was also examined. This was followed by a CFA of the self-rated proficiency measures to determine whether the items measuring proficiency in these different varieties did in fact load onto the three hypothesised latent variety factors. Reliability analyses were also conducted to determine the scores for each item and the legitimacy of retaining them in subsequent analyses.

¹⁶ Henceforth referred to as either Overall higher-order construct model or simply 'Overall'.

4.1.1. Second Language Motivational Self-System

4.1.1.1. Preliminary Analyses

4.1.1.1.1. *Analysis of Missing Data*

The computation of factor scores using the R function *LavPredict* (Rosseel, 2012) requires complete entries for all participants. Two options were thus available: remove cases with any missing entries, or enter imputed values for missing responses.

All Likert-scale items for the motivation and self-rated proficiency questions at T1 and T2 were entered into SPSS (version 25), and the data was analysed to first determine the extent of any missing values.

4.1.1.1.1.1. *Motivation Scales*

The pattern of missing data was examined to determine whether the arrangement of missing values was random. Since missing values were not confined to particular questions, and thus randomly distributed, the dataset was considered suitable for imputation.

An item-by-item analysis indicates that between 2.2 and 5.2% of responses were missing for each variable at T1. This number was higher at T2, even after the participants from schools who did not participate at this timepoint were removed (referred to subsequently as the 'adjusted sample'). In this adjusted sample at T2, items were missing between 19.1 and 23.7% of responses. The full breakdown of missing items can be found in Appendix C.1.

The charts for missing items for T1 (Figure 4.2.) indicate that removing all *cases* (i.e. participants) with any missing entries in these items would result in a loss of 17.5% of the dataset. To avoid such a substantial reduction of the dataset, missing *values* (i.e. responses), which constituted 3.5% of the overall data, were imputed instead. This also facilitated retention of a greater proportion of the dataset at T2, since if participants with missing data were to be removed the dataset would be cut by almost a third.

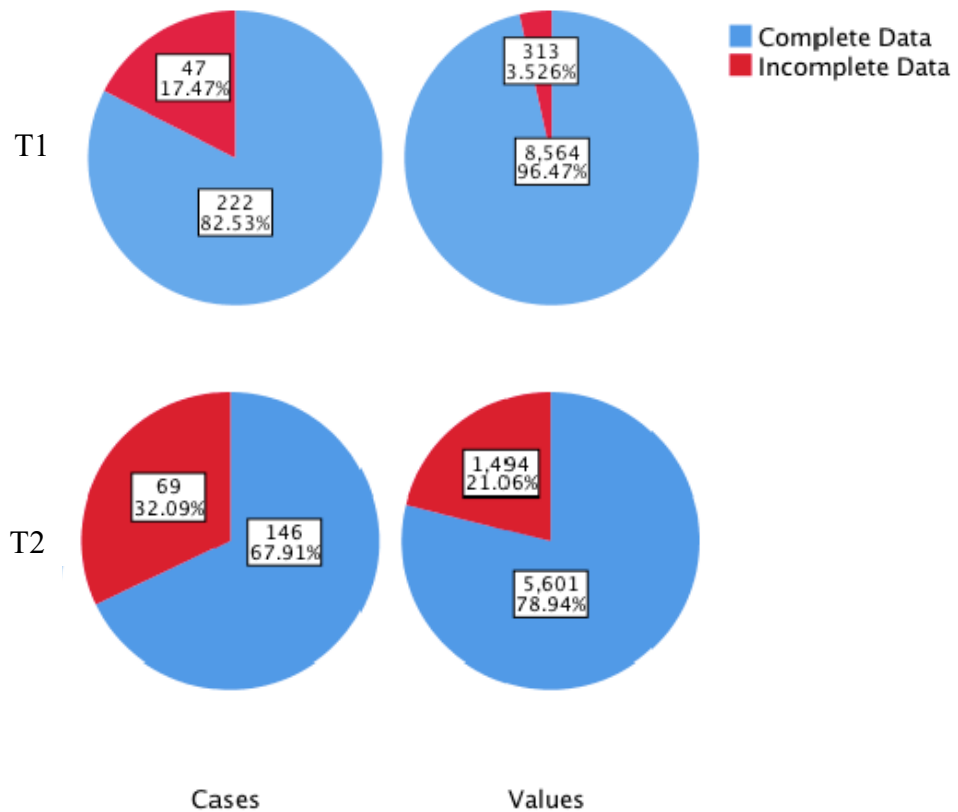


Figure 4.2. Missing Data Summary (Motivation Scales T1 & T2)

4.1.1.1.2. Self-Rated Proficiency

Similar proportions of missing data were observed in the Self-Rated Proficiency scales, with 1.9 to 3.3% of data missing per item at T1, and 20 and 21.9% in the adjusted sample at T2. Details of the missing data for each item at both timepoints, as well as for the non-adjusted T2 sample, can be found in Appendix C.2.

An analysis of missing data in SPSS indicated that removing cases with any missing responses at T1 would only result in a loss of 6.7% of the data (Figure 4.3). This, however, would be a far more substantial loss at T2, with over a quarter of the data being deleted. Thus, again, missing values were imputed.

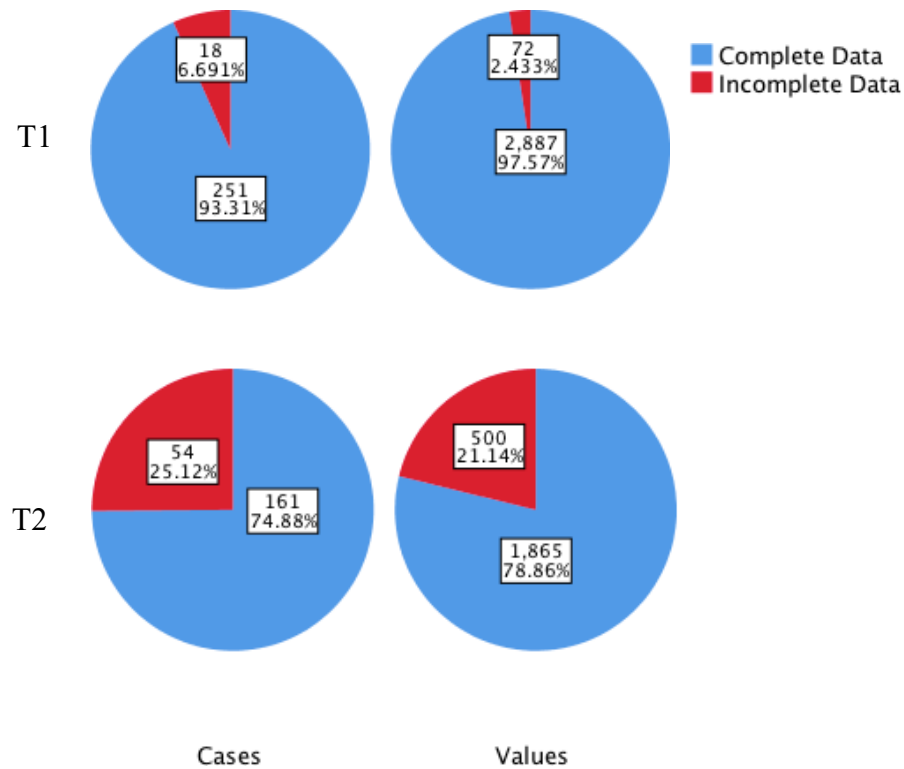


Figure 4.3. Missing Data Summary (Self Rated Proficiency Scales T1 & T2)

4.1.1.1.2. Treatment of Missing Data

The data was transferred to R to impute missing values. A multiple imputation method was run using the ‘mice’ package (van Buuren & Groothuis-Oudshoorn, 2011). This was considered preferable to replacing missing values with the series mean, which reduces the variance of the dataset.

The ‘mice’ package runs iterations using the existing data to find a set of replacement values that best fits the data that is present for participants with usually <5% of their entries missing (Alice, 2015). Since the initial analysis revealed that two participants had just over 5% missing, and there is some flexibility in the threshold (Madley-Dowd, Hughes, Tilling et al., 2019) the limit was set at 6%. Participants missing more than this limit were removed from the dataset, leaving the totals specified in Table 4.1. In other words, only responses for participants with <6% of values missing from their responses were imputed. Participants with >6% of values missing were excluded from the dataset. Once the analysis was run, one dataset with completed missing entries was returned.

This dataset is used throughout for all subsequent analyses involving these items.

Table 4.1: Motivation Scales—Missing Data Summary					
	No. of Cases	Proportion of Total Cases at T1	No. of valid cases following imputation	Imputed cases as a proportion of total at T1	Proportion of original total no. of cases
T1	265	100%	250	100%	94.3%
T2	178	67.2%	152	60.8%	57.4%

4.1.1.2. Normality and Outliers

The data was first tested for normality of distribution and outlying values to determine both the suitability of the data for running factor analyses and to determine which extraction and rotation methods to employ.

Outliers were identified by transforming the item scores into z-scores. A common criterion for identifying outliers is that they are >3.29 standard deviations removed from the sample mean (Field, 2018). Based on this criterion, no outliers were identified in the motivation scales data at either Time 1 or Time 2.

Checks for normality and outliers were carried out in SPSS for all three sets of items at both timepoints. First, the data was examined visually: Q-Q plots and histograms were generated for each of the items at both timepoints. These were visually inspected to check for the shape distribution and position of the best fit curve. For the majority of items, the data visually aligned with the expectations of normality.

Normality of the item scores was also tested through the calculation of z-scores for skewness and kurtosis and by dividing these raw values by their respective standard error. This procedure is more robust and avoids the subjectivity of visual inspection of normality by providing clear cut-off points. In large samples such as that at T1, z-scores above 3.29 are commonly taken to indicate non-normality (Field, 2018).

At T1, all the Current Self (CS) items were normally distributed across the language varieties, with skewness z-score values ranging from $z=-1.69$ to 2.73 . For the Ideal Self

(IS), skewness z-scores ranged from $z=-5.74$ to -0.64 . Of the four items, only IS Desire was significantly negatively skewed across the language varieties with MSA IS Frequency and Vividness scores showing negative kurtosis, which ranged across the items from $z=-4.00$ to -0.45 . The Ought-to Self (OTS) items diverged from normality most noticeably. The scores for OTS Force were positively skewed across all three varieties, along with some significant skewness and kurtosis in a number of other items across one or two varieties.¹⁷ These skewness scores ranged from $z=-4.79$ to 9.55 , with the kurtosis values ranging from $z=-3.92$ to 4.76 .

At T2, similar patterns were observed. For CS items, skewness and kurtosis z-scores lay within the acceptable levels, ranging from $z=-1.05$ to 2.26 , and $z=-2.84$ to -1.20 respectively. Skewness values for the IS items were also all within the acceptable range with the exception of Classical Desire ($z=-4.00$) and Dialect Desire ($z=-4.31$). The kurtosis scores for these items were also generally within range ($z=-3.86$ to -1.32), with a number of items falling beyond the 3.29 cut off. The full table of descriptive statistics can be found in Appendix C.3. Again, OTS items showed a strong positive skew ($z=-0.30$ to 7.26), while the kurtosis scores were mostly within range ($z=-3.85$ to 1.91).

Based on an interpretation of the indicators above, normality can be assumed for the majority of items, yet a smaller number of scores were non-normally distributed. As a result, robust estimation methods were employed during later CFAs, namely robust maximum likelihood estimation (Blunch, 2013; Fabrigar, Wegener, MacCallum, et al., 1999).

4.1.1.3. Correlations

It is theorised that the items for each dimension of learners' Motivational Self concepts will be correlated across language varieties, as well as across timepoints. The relationship between these constructs (CS, IS, OTS) requires examination in order to

¹⁷ Kolmogorov-Smirnov and Shapiro-Wilk tests were also run for each set of items at T1 and T2. Scores for each of the items were significant ($p<.005$), meaning that the structure of the data differs significantly from a normal distribution. Although this test indicates that the data violates the assumption of normality, it is worth noting that this method is vulnerable to size bias (Field, 2018) and is therefore used merely as confirmation of findings from the z-score examinations.

determine whether or not this is in fact the case and, if so, what the degree of correlation between them is. Bivariate, non-parametric, two-tailed correlation analyses were run between the self-concept scale items at T1, and at T2 separately. Since some of the items violated parametric assumptions of normality, Spearman's rho was used (Field, 2018).

There were statistically significant positive intra-variety correlations between the items of each self-concept (CS, IS, OTS), as well as with their corresponding constructs across varieties (MSA, Classical, Dialect), without exception. In addition, there were highly significant positive correlations between CS and IS items across varieties at both timepoints, with the exception of only the following at T1: *Classical IS Desire* and *Dialect CS Self-Perception* ($r_s=.110$); *Dialect IS Desire* and *Classical Self-Perception* ($r_s=.081$); and *MSA CS Ability* and *Classical IS Desire* ($r_s=.105$). All correlation tables can be found in Appendix C.4.

Intra- and inter-variety correlations between OTS and IS/CS items tended to be weaker or non-significant. One exception to this is the *OTS Social Pressure* item which correlated with all CS items at T1 across and within varieties. This item also correlated with 5 of the 9 CS items across all language varieties at T2.

Correlation coefficients were also examined between items at T1 and T2. Spearman's rho was used in conducting a bivariate, non-parametric test. On this occasion, one-tailed correlations were run, since T1 is hypothesised to predict T2. Similar patterns were observed: intra-variety items correlated with each other for CS and IS, and for the most part for OTS. Inter-variety items at T1 also correlated with their corresponding items at T2 for CS and IS, and for the most part for OTS.

CS items at T1 correlated with IS items at T2 across language varieties, and vice versa, with only three exceptions: *Dialect IS Desire* with *Classical CS Ability*; *Dialect IS Desire* with *CS Self-Perception*; and *MSA IS Desire* with *Classical CS Self-Perception*. Again, far fewer correlations were observed between OTS and CS/IS items between timepoints, suggesting that the constructs may only be loosely related.

It is evident from these correlation statistics that all variables had at least one

correlation coefficient greater than 0.3, and that there were strong correlations between multiple items for all language varieties. This made the dataset suitable for subsequent factor analyses. These preliminary analyses also suggest that the OTS may not be as strongly related to the CS as its other future-self counterpart, nor in fact closely related to the IS. This corroborates the findings of other research studies, such as Al-Hoorie's (2018) meta-analysis, which demonstrated the differing correlations between the IS, OTS, and outcome variables.

4.1.1.4. Analysing the Structural Validity of the L2MSS

To determine the underlying structure of the L2MSS, as measured by the Motivation Survey scales, a CFA was conducted. Specifically, this aimed to explore whether the *a priori* factor structure, constituted of CS, IS, and OTS factors per Arabic variety, fits the data.

4.1.1.4.1. *Confirmatory Factor Analysis*

CFA is a type of structural equation modelling (SEM) which examines the relationship between observed or measured variables (indicators) and underlying latent variables (factors) to determine the validity of the underlying structure of a construct where an *a priori* model exists (Brown & Moore, 2012). CFAs were run in R using the lavaan package (version 0.6-3; Rosseel, 2012) and graphed using semPlot (version 1.1; Epskamp, 2017).

Due to the presence of some non-normally distributed variables, robust maximum likelihood (MLR) estimation was used. As is common practice, the factor loading of the first item measuring each latent construct was fixed to one to allow for model identification (e.g., Blunch, 2013; Byrne, 2010; Kline, 2015).

The robust versions of four goodness-of-fit measures were examined: (1) the Comparative Fit Index (CFI) and (2) the Tucker Lewis Index (TLI), for which a value of >.9 is considered acceptable, and >.95 excellent; and (3) the Root Mean Square Error of Approximation (RMSEA), and (4) Standardized Root Mean Square Residual (SRMR), for which acceptable scores are <.1, and preferably <.05 (Hopwood & Donellan, 2010; see

Hu & Bentler, 1999 for more conservative estimates). Model Chi-square (χ^2) scores were also noted, with lower values indicating comparatively better fits between models. Ideally, chi-square values should not reach significance for a model to be deemed a good fit. However, given the sample size in this study, the chi-square values usually reached significance, and so the p -values were not considered when determining goodness-of-fit (Hu & Bentler, 1999).

4.1.1.4.2. *Time 1*

For the MSA items, the initial goodness-of-fit statistics for the theorised model, where each item loaded onto its respective latent self-concept factor, were outside the acceptable thresholds. The scores were improved by removing the IS Plausibility item, although the TLI (.883) and RMSEA (.115) remained outside acceptable bounds. Two further modifications were made to the model, where IS Frequency and Desire, as well as CS Ability and Confidence, were correlated in order to control for the covariances between the two sets of items. Upon doing so, the goodness-of-fit measures produced were within the recommended ranges. These measures are provided in full in Appendix C.5.

For Dialect at T1, the inclusion of IS Plausibility did produce acceptable goodness-of-fit scores. However, the model fit indices show that not including the Plausibility item, as had been done for the MSA model, produced better scores. These scores were further improved by again correlating IS Frequency and Desire.

For the Classical variety, the fit indices indicated that including IS Plausibility as loading onto CS produced a less satisfactory fit than excluding it or than including it as an IS item. Once again, excluding the item, plus controlling for the covariances between IS Frequency and Desire produced the best model fit scores. Factor loadings for each of the items can be found in Table 4.2 (items with corresponding colours were correlated to control for their covariances).

Table 4.2: CFA Factor Loadings, Errors, and Item Covariances T1									
Variable	MSA (Fit 3)			Dialect (Fit 2)			Classical (Fit 2)		
	Factor Loading	Error	Covariance	Factor Loading	Error	Covariance	Factor Loading	Error	Covariance
<i>CS Ability</i>	.76	.42		.85	.27		.85	.28	
<i>CS Confidence</i>	.97	.05	-2.13	.83	.32		.80	.37	
<i>CS Self-concept</i>	.88	.22		.71	.50		.67	.55	
<i>IS Desire</i>	.66	.56	.33	.60	.64	.35	.55	.70	.41
<i>IS Frequency</i>	.78	.40		.74	.45		.66	.56	
<i>IS Vividness</i>	.96	.07		.95	.10		.93	.14	
<i>OTS Consequences</i>	.74	.45		.70	.52		.62	.62	
<i>OTS Disappointment</i>	.87	.25		.93	.14		.93	.13	
<i>OTS Force</i>	.63	.60		.65	.58		.53	.72	
<i>OTS Social Pressure</i>	.60	.64		.54	.71		.44	.80	

After IS Plausibility was removed and once pairs of items suggested in the modification indices were correlated, the goodness-of-fit measures, and the individual item factor loadings themselves, confirmed the validity of the hypothesised three-factor model at T1.

It appeared from the goodness-of-fit indices at T1, particularly for Classical and Dialect, that loading Plausibility onto the CS produced acceptable measures. Whilst these were improved further by removing the item altogether, the initial fit indices suggest that Plausibility relates not only to the IS, as hypothesised, but also to the CS, tapping into a potential underlying association between the two. Given the fact that the likelihood of achieving a future self-concept, which is what the Plausibility item addressed, is partially dependent on an individual's current state, the inference of a cross-loading of this item onto CS and IS is not unexpected. Since the fit indices were improved by removing, rather than cross-loading this item, Plausibility scores were thus reserved for subsequent analyses involving the concept of Self-Discrepancy—the gap between the IS and the CS (see Section 4.5).

4.1.1.4.3. Time 2

Since Plausibility had been removed from each set of items at T1, the decision was made to remove it from model fit analyses at T2 also and to analyse its relationship with other constructs in the study independently.

A similar model emerged at T2 for the MSA items, where (a) the removal of IS Plausibility and (b) controlling for covariances between IS Frequency and Desire, and between CS Ability and Confidence, produced better fit indices than the original theorised model. The RMSEA for this particular model was .102, being just outside the acceptable range. This was remedied by further modifying the model to control for covariances between OTS Consequences and Social Pressure (see Appendix C.5 for full fit measures).

For Dialect and Classical Arabic at T2, the hypothesised model with Plausibility removed again produced acceptable fit indices, which were further improved by controlling for correlations between IS Frequency and Desire, and CS Ability and Confidence respectively. The resulting factor loadings can be seen in Table 4.3 (items with corresponding colours are correlated in the model to control for their covariances).

Table 4.3: CFA Factor Loadings, Errors, and Item Covariances T2									
	MSA (Fit 3)			Dialect (Fit 2)			Classical (Fit 2)		
Variable	Factor Loading	Error	Covariance	Factor Loading	Error	Covariance	Factor Loading	Error	Covariance
<i>CS Ability</i>	.75	.44		.88	.22		.74	.46	.51
<i>CS Confidence</i>	.77	.41	.56	.88	.22		.77	.40	
<i>CS Self-concept</i>	.89	.21		.78	.39		.77	.41	
<i>IS Desire</i>	.69	.52	.47	.67	.56	.42	.81	.34	
<i>IS Frequency</i>	.77	.40		.80	.36		.90	.19	
<i>IS Vividness</i>	.95	.09		1.00	0		.82	.33	
<i>OTS Consequences</i>	.86	.26	-.58	.82	.33		.62	.61	
<i>OTS Disappointment</i>	.79	.38		.90	.19		.77	.41	
<i>OTS Force</i>	.66	.57		.63	.61		.46	.79	
<i>OTS Social Pressure</i>	.75	.44		.53	.71		.63	.60	

4.1.1.4.4. *Three Factor Structure*

Once model fits had been examined for each language variety independently at T1 and T2, a model combining all varieties, and correlating the items across the varieties, was tested in order to examine whether this three-factor structure was a suitable fit for the data as a whole at each timepoint.

CFI, TLI, RMSEA, and SRMR measures for all three varieties combined into a single model at both T1 and T2 indicated that the three-factor structure, incorporating CS, IS, and OTS, was indeed a good fit for the data (see Appendix C.6).

As well as offering a method via which to evaluate an instrument's validity, CFAs are also used to produce *factor scores*. These values represent a numerical sum of the individual item scores weighted according to each item's factor loadings. These were generated using the `lavPredict` function in R for the three-factor structure models, to be used in subsequent path model analyses to represent the latent constructs (DiStefano, Zhu, & Mîndrilă, 2009).

As expected, the indicators all revealed significant positive factor loadings, with standardized coefficients ranging from .46 to .93 (see Figure 4.4). The latent variables for each self-concept for each language variety (e.g., MSA CS, Dialect CS, Classical CS etc.) also loaded onto a single higher-order construct ('Overall' CS; IS; OTS) as illustrated in Figure 4.4. The loadings of each variety onto the higher-order latent factor are represented in bold in Appendix C.7.

There were also positive correlations among all three latent factors (see Appendix C.7), indicating that students who recorded stronger CS constructs were more likely to possess stronger IS and OTS constructs and vice versa.

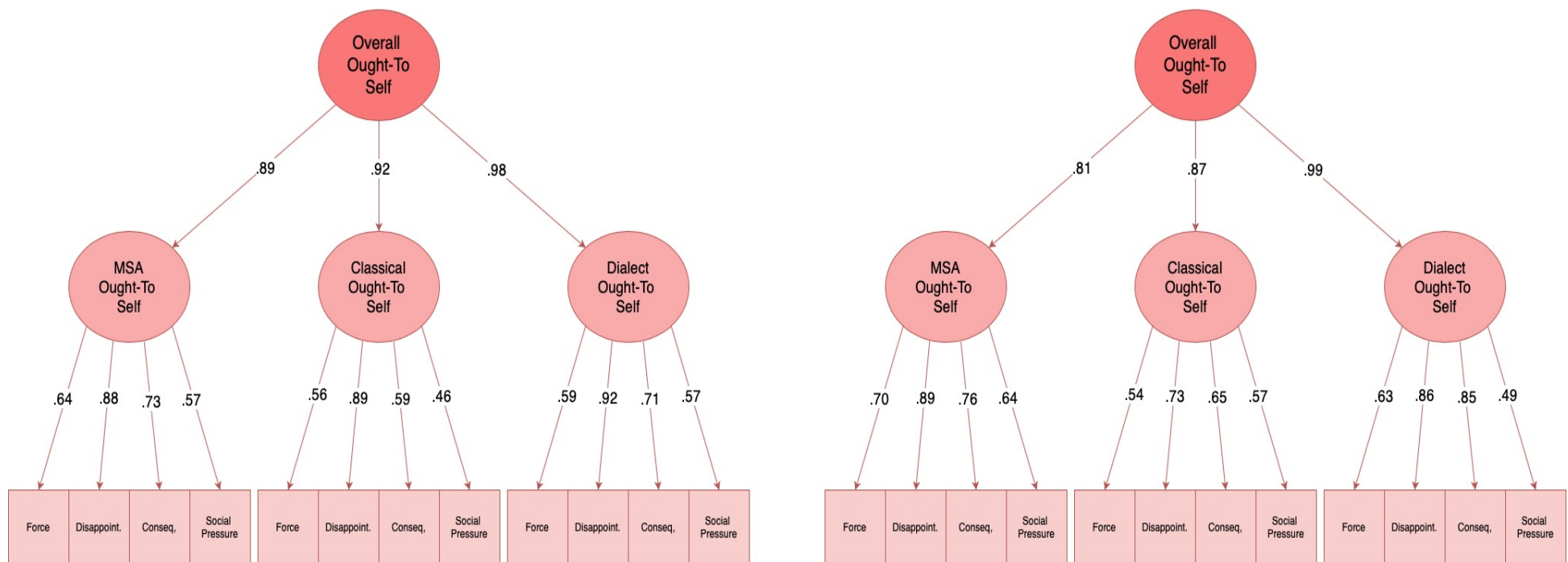
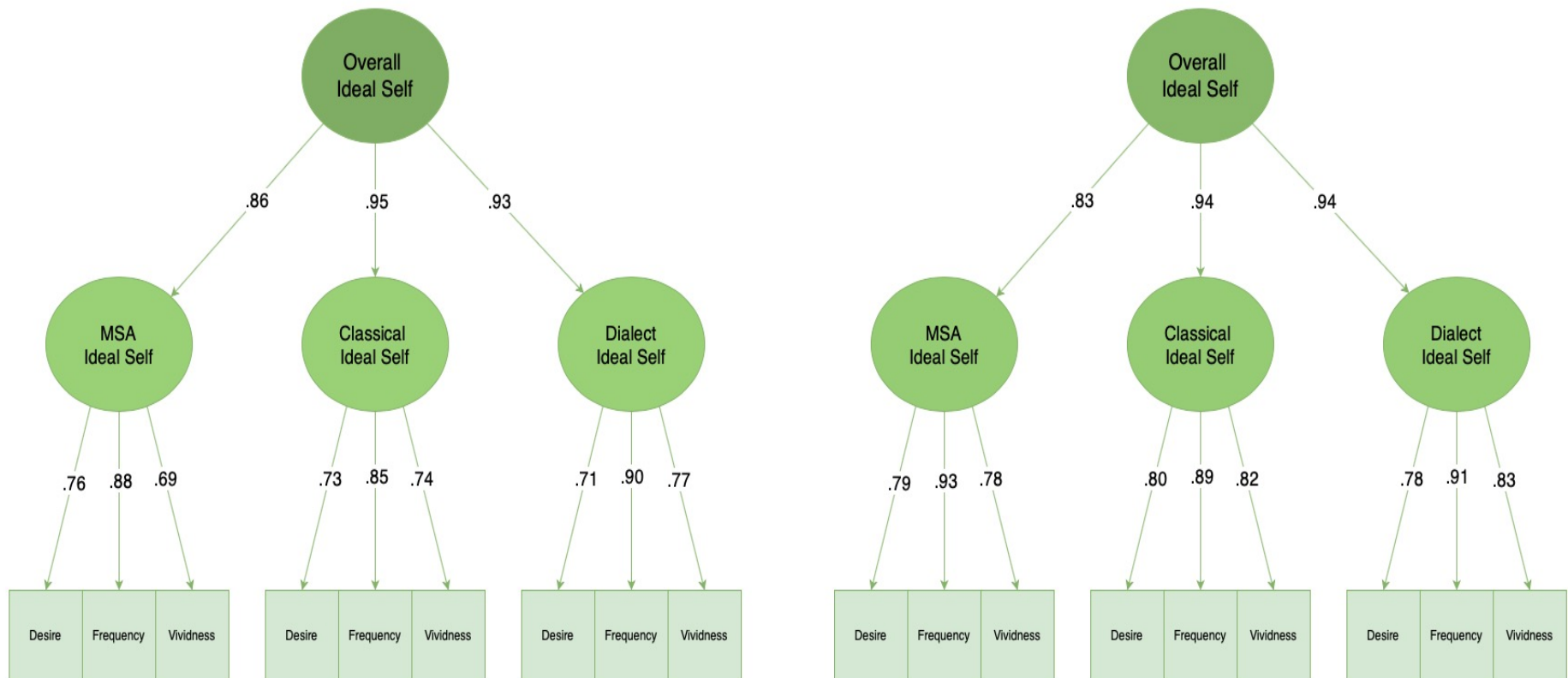
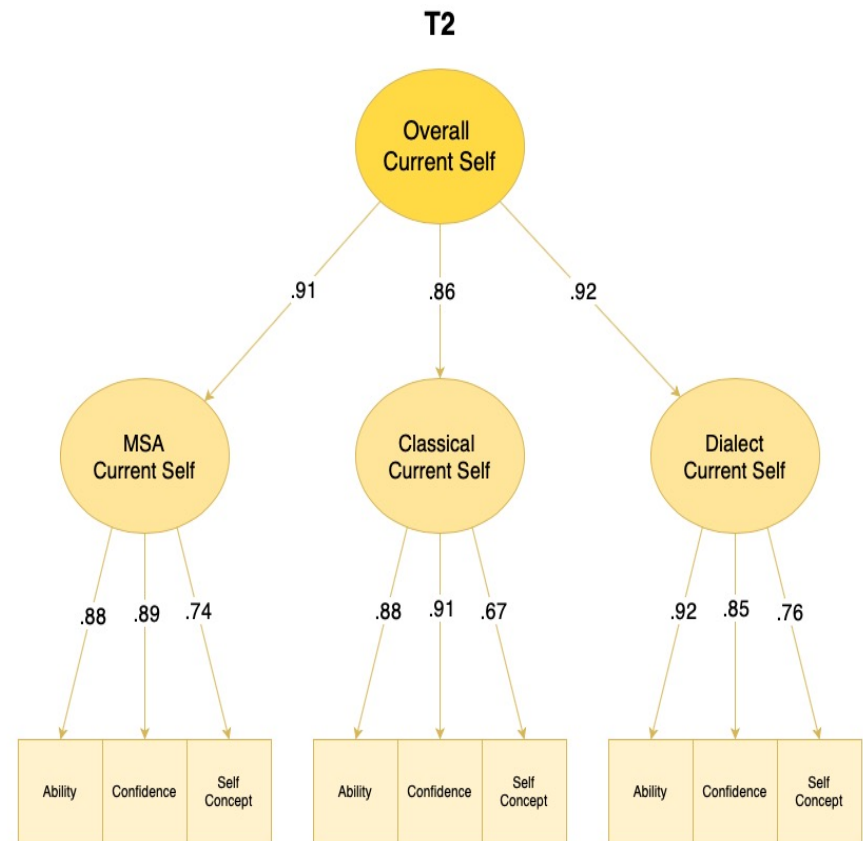
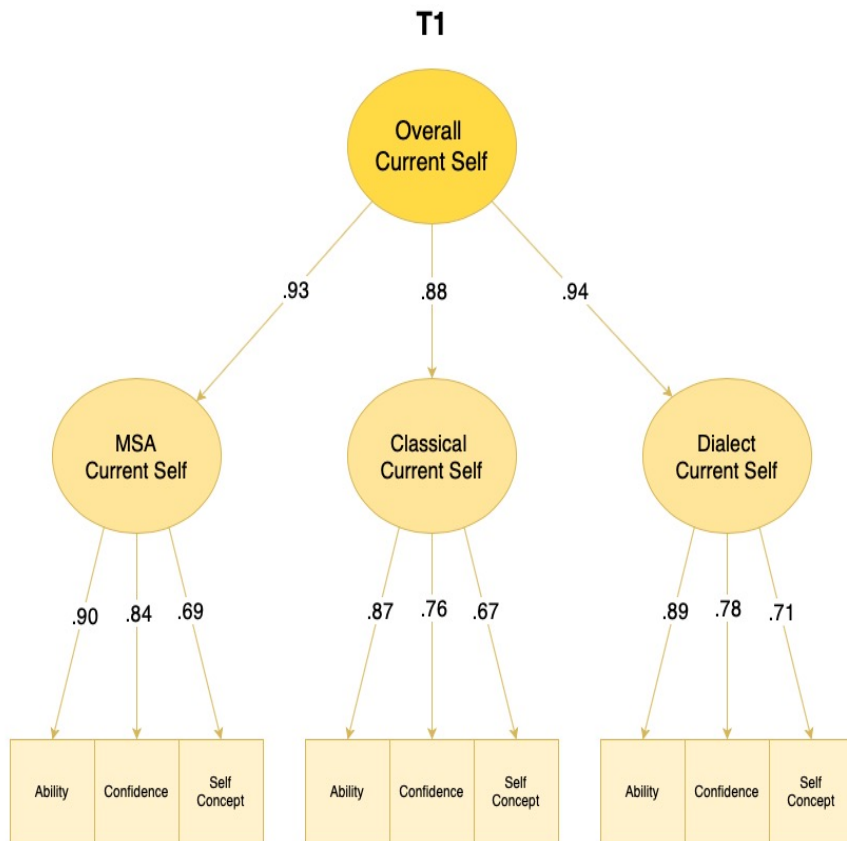


Figure 4.4. Three-Factor Structure CFA Diagrams (OTS – IS – CS)
Squares represent items; circles represent latent variables; factor loadings recorded on arrows.





4.1.1.5. Instrument Measurement Quality

Whilst the CFA offers strong support for the underlying structural validity of the L2MSS as measured by the motivation scales, it is important to assess the reliability of this instrument in order to determine the quality of the measurements provided. This was done by examining the reliability scores for each observed variable.

4.1.1.5.1. *Reliability: Cronbach's Alpha*

An analysis of Cronbach's Alpha scores for each of the self-concept scale items indicated that the CS, IS, and OTS measures had good internal consistency at both timepoints for all language varieties (Table 4.4). All scores fell within the range of $\alpha=.71$ to $.89$ at T1, with the majority over $\alpha=.8$, and at T2 from $\alpha=.71$ to $.90$. Item-total correlations showed that all items correlated sufficiently with the scale as a whole ($r>.3$; see Field, 2018), the lowest of which was the Classical OTS Social Pressure measure at T1 ($r=.386$).

For 6 of the 33 items at both timepoints the adjusted alpha scores indicated that there would be a slight improvement in the overall consistency of the measure if these were removed. However, since none of these items would lead to an increase in the overall alpha score of more than 0.034 if removed, and the initial overall alpha scores were all above 0.7, the scales were considered appropriate, reliable, and valid measures of participants' L2 self-concepts. A full table of the reliability scores per item can be found in Appendix C.8.

Table 4.4: Motivation Scale Reliability Scores (T1 & T2)							
Construct	Variety	N	No. of Items	Cronbach's Alpha	Cronbach's Alpha (Standardised Items)	Suggested delete item	Cronbach's Alpha if deleted
T1							
IS	Classical	260	4	.797	0.796	Plausibility	0.81
	Dialect	261	4	.847	0.846	-	-
	MSA	259	4	.886	0.886	-	-
CS	Classical	259	3	.802	0.806	Current SC	0.819
	Dialect	260	3	.831	0.834	-	-
	MSA	256	3	.862	0.864	Current SC	0.864
OTS	Classical	259	4	.71	0.708	Soc Pressure	0.713
	Dialect	260	4	.782	0.782	Soc Pressure	0.797
	MSA	257	4	.791	0.798	Soc Pressure	0.797
T2							
IS	Classical	172	4	.890	0.89	-	-
	Dialect	169	4	.889	0.889	Plausibility	0.89
	MSA	166	4	.904	0.904	-	-
CS	Classical	172	3	.850	0.851	Current SC	0.884
	Dialect	170	3	.877	0.88	Current SC	0.884
	MSA	167	3	.879	0.88	Current SC	0.887
OTS	Classical	171	4	.713	0.712	-	-
	Dialect	170	4	.784	0.786	Soc Pressure	0.816
	MSA	166	4	.819	0.828	Soc Pressure	0.826

4.1.1.6. Summary

The positive correlations between the scale items and the strong factor loadings resulting from the CFA offer support for the underlying validity of the L2MSS structure and its three constituent self-concepts: CS, IS and OTS. These distinct latent variables emerged for all three varieties, suggesting the construct's validity when applied to different language types. Though a three-factor model was supported, the fact that factor loadings differed for each individual variety indicates the validity of viewing Arabic as a multiglossic phenomenon. It also highlights the importance of examining the nature of learners' motivational selves in relation to each one so as not to overlook important nuances in interactions with these three varieties.

The best model emerged once the IS Plausibility item was removed. The issue of the joint loading of these IS Plausibility items in different language varieties is noteworthy, since it hints at the underlying mechanism of the original theory, which suggests that motivation is derived from the gap between the current and future selves and the desire to decrease it. Since plausibility considers the likelihood of achieving a future self-image, which will depend on the distance between the CS and IS, it is hypothesised that this will correlate with the size of the gap between learners' self-reported CS and IS concepts. In other words, the discrepancy between these constructs, or self-discrepancy (SD), is posited to constitute one of the major mechanisms underpinning the L2MSS. There is presumed to be a degree of elasticity in S-D and its correlation with motivation: beyond a certain point, the possibility of overcoming the distance between the CS and IS becomes implausible and thus demotivating. These hypotheses and relationships are examined in response to RQ 5 (Section 4.5).

In addition to offering support for the structural validity of the underlying constructs, the analysis of the internal consistency of the instruments used to measure these yields positive results. The reliability scores for the items at both timepoints were well within the acceptable range. The suggestion of minor improvements resulting from the deletion of items were not considered significant enough to warrant their removal, and given the fact that there were limited items for each construct to begin with (a decision made so as to avoid fatigue effects amongst participants), all items were retained.

4.1.2. Self-Rated Proficiency

Participants were asked to report their perceptions of their current proficiency levels at both timepoints in skills related to the different language varieties. Since this scale is unique and previously untested, the same set of analyses were run on the data to determine the structural validity of the underlying proficiency construct, and to validate the scales themselves. An overview of the missing and imputed data totals for Self-Rated (SR) Proficiency scales can be found in Section 4.1.1.1.1.

4.1.2.1. Preliminary Analyses

4.1.2.1.1. *Normality and Outliers*

In order to examine the relationship between SR Proficiency and learner's motivational selves, as well as to determine whether the Likert-scale items were reliable measures of this construct, the same methods as the Motivation Survey scales were repeated, with the imputed dataset.

Checks for normality and outliers were carried out in SPSS. No outliers were identified in this dataset at either Time 1 or Time 2, ascertained by an examination of the z-scores for these items per participant. A visual inspection of the Q-Q plots and histograms, as well as examination of the z-scores for skewness and kurtosis, reveals that the majority of self-reported proficiency ratings do not meet assumptions of normality, either exhibiting significant skewness, kurtosis or both, at one or both timepoints; the z-scores for skewness ranged from -19.62 to 1.61, and those for kurtosis from -4.32 to 29.09. A full account of the descriptive statistics can be found in Appendix C.9. As a result of these non-normal distributions, non-parametric tests and robust statistical measures were employed in further analyses.

4.1.2.1.2. *Correlations*

Items were intended to measure learners' proficiency in skills related to each of the language varieties (Modern Standard, Classical, and Dialect). Bivariate, non-parametric, two-tailed correlation analyses using Spearman's rho were run between items at T1 and at T2 separately, followed by a one-tailed analysis comparing items at T1 and T2 to examine the correlations between items.

The majority of items exhibited positive correlations with each other within both timepoints, with strong correlations exhibited between items within each variety (see Appendix C.10). Statistically significant correlation coefficients ranged widely from .129 to .796 at T1, and .153 to .892 at T2, with most in the .3 to .5 range. The only items which appear not to correlate significantly are those involving word reading (with and without diacritics) and Dialect speaking and recognition/comprehension. This is not surprising

given that dialects are mostly oral and rarely written down. The same items at T1 are also uncorrelated with the other at T2. In addition, MSA Writing and Dialect Speaking skills in which are arguably only considered loosely related (Albirini, 2014) are also uncorrelated between T1 and T2.

4.1.2.2. Structural Validity of the SR Proficiency Scales

The individual proficiency-related scale items were intended to measure the level of learners' abilities across the three Arabic language varieties. A confirmatory factor analysis was performed on the data to examine whether or not this theorised three-component structure was present in the data.

4.1.2.2.1. Confirmatory Factor Analysis

To test the validity of the self-rated proficiency measures, a CFA was again run in R using the lavaan package (Rosseel, 2012) for data at both timepoints. Robust maximum likelihood (MLR) estimation was used given the presence of some non-normally distributed variables, and factor loadings of the first item measuring each latent construct was again fixed to one to allow for model identification (e.g., Blunch, 2013; Byrne, 2010; Kline, 2015).

Multiple hypothesised models were explored. The original three-factor structure with *Word Reading with Harakat* loading onto a latent Classical proficiency factor was first trialled (full fit indices are provided in Appendix C.11). The indices for this original hypothesised model with three latent factors and all items included were, for the most part, acceptable, but not ideal, with RMSEA being above the .1 threshold (RMSEA=.107). The *Word Reading with Harakat* item loaded only weakly onto the Classical latent variable (0.27, $p=.001$). Whilst the *Word Reading* item loaded onto MSA, its loadings (0.58, $p<.001$) stood out as lower than other items it clustered with, the loadings of which ranged from .80 and .86 ($p<.001$).

The modification fit indices for this model suggested that it would be improved by loading the *Word Reading with Harakat* item onto the MSA latent variable ($mi=33.695$,

epc=0.491), and by correlating this item with the *Word Reading* item ($mi=31.079$, $epc=0.256$). It was subsequently hypothesised that these two items may measure a separate underlying proficiency in phonetic decoding and letter recognition independent of any language variety. A model in which these two loaded onto a 'Basic Literacy Skills' latent variable was trialled. This produced superior fit indices, and an improvement in the Chi square model fit value ($X^2=96.524$, $p<.001$), though this still reached significance as a result of the limited sample size.

These items were considered to provide little useful information with which to discriminate between students for two reasons. Firstly, a re-examination of the descriptive statistics indicated that most students scored highly on these items ($M=5.64$, $SD=.969$), which is to be expected since the majority of students were from a Muslim background in which it is traditional to learn to decode the Arabic script for religious practices, and since most students indicated they had completed at least 3 years of Arabic instruction (62.9% at T1). And secondly, these items appeared to cross-load onto multiple latent variables. The original 3-factor model with these items removed was analysed. This produced acceptable model fit indices, though largely similar to the previous model; however, the Chi square was almost halved ($X^2=54.555$, $p<.001$), indicating that the model without these items was, indeed, a better fit than the previous two models. Modification indices were again examined, and the cross-loading the MSA v. Dialect item ($mi=12.743$, $epc=0.438$) was implemented. Though the loading of the item onto the MSA factor is low, this solution was reasonable, since the ability to distinguish between Dialect and Standard Arabic requires a familiarity with both these varieties, and the resulting fit indices imply the modified model is an improved match for the data.

Similar patterns were observed when these models were replicated with T2 data. Model 3a produced the lowest Chi square value ($X^2=59.75$, $p<.001$), though this did reach significance as a result of the limited sample size. The only fit measure outside of the generally accepted limits was the RMSEA, which was .01 above the acceptable range. Given the acceptable values of the other fit indicators and the suitability of this model for T1 data, it was determined that this model would be adopted for both times points and used to compute factor scores, with the factor loadings produced and recorded in Table 4.5.

Table 4.5: CFA Factor Loadings—Self-Rated Proficiency T1&2												
Variables	T1						T2					
	MSA		Dialect		Classical		MSA		Dialect		Classical	
	Loading	Error	Loading	Error	Loading	Error	Loading	Error	Loading	Error	Loading	Error
Self-Rated Proficiency	.73	.47	.70	.55	.68	.54	.59	.65	.68	.57	.83	.31
MSA Writing	.74	.45					.84	.29				
MSA Speaking	.84	.29					.86	.27				
MSA Reading Comprehension	.89	.20					.91	.16				
MSA Listening Comprehension	.75	.44					.88	.23				
Distinguishing MSA & Dialect	.28	.55	.51	.55			.29	.57	.50	.57		
Dialect Speaking			.81	.34					.89	.21		
Dialect Listening Comprehension			.92	.15					.89	.20		
Quran Reading Comprehension					.90	.19					.95	.09
Quran Listening Comprehension					.88	.23					.92	.15

4.1.2.3. Instrument Measurement Quality

As with the motivation scales, it is important to assess not only whether the theorized underlying structure of the SR Proficiency construct exists, but whether the instrument is a reliable and valid measure of this to begin with. This was done by running reliability analyses for each of the scale items.

4.1.2.3.1. Reliability

The Cronbach's alpha scores for each item were examined at each timepoint (see Appendix C.12). The scale overall displayed strong reliability scores at both T1 ($\alpha=.867$) and T2 ($\alpha=.897$), and the scores for individual items suggested this would be improved only by deleting one item *Word Reading with Harakat* at T1 ($\alpha=.869$). Although this was low, this does corroborate the findings of the CFA which suggest that the word reading items do not fit the latent variable well. These items were removed from further analyses.

4.1.3. Relationship between Proficiency Tests and Self-Rated Proficiency

The validity of the SR Proficiency measures was also examined through a comparison with scores for the two proficiency tests: Yes/No and cloze tests at T1 and T2.

4.1.3.1. Proficiency Test Descriptive Analyses

There were no outliers for either the Y/N or cloze tests at T1 or T2. Whilst the skewness z-scores for the Y/N total at T1 were both slightly above the 3.29 cut off, the raw scores were well within the acceptable limits for parametric tests, which were subsequently employed (see Appendix C.13 for full descriptive statistics).

4.1.3.2. Correlations

Although the Proficiency test scores met the assumptions for normality, Spearman's rho correlation values were examined for the self-rated and proficiency test scores since the former did not meet assumptions of normally distributed data. Correlation coefficients are reported in Appendix C.14.

At T1, the scores for the Y/N and cloze tests correlate significantly with the SR Proficiency scores for each variety, though most strongly with MSA. This is expected given that the proficiency tests involve materials derived from the GCSE Specification which employs only MSA. The coefficients for the Y/N tests are higher than those for the cloze tests. This is also expected as the Y/N tests measured learners' vocabulary knowledge, some of which is transferable across varieties. The cloze tests on the other hand, presented learners with full sentences and tapped into grammatical knowledge which differs between standard and colloquial Arabic. This would explain why the correlations between cloze tests and Dialect SR Proficiency were the lowest. Coefficients were lower at T2, though the test and self-rated measures were still significantly correlated. It is also likely that the reason these correlation coefficients were not higher is that the SR Proficiency scales tapped into all four skills, whereas the tests only focused on reading. Nonetheless, the existence of significant correlations between the

subjective and objective measures adds support to the validity of the SR Proficiency scores which tap into an underlying proficiency construct.

4.1.4. Summary

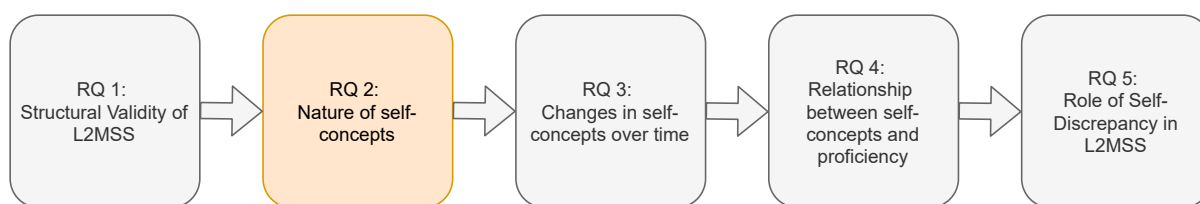
The above analyses suggest good overall validity and reliability of the self-rated proficiency scale when the items pertaining to word reading are removed. These items did not load strongly onto any one of the variables. The lack of significant factor loadings is not unexpected, since word reading is not variety-specific. In addition, the strong negative skewness suggests the redundancy of these items, since all learners taking the GCSE, as expected, were able to decode words with or without diacritics. It is thus considered acceptable to remove these items from further analyses. The remainder of the items exhibit strong reliability scores; thus, the instrument can be considered to be a valid measure of learners' self-perceived multiglossic proficiency.

The nine items retained from the scales loaded effectively onto the three underlying language variety variables, offering support for a model of proficiency which distinguishes between the underlying tri-partite structure of the Arabic language, consisting of MSA, Dialect, and Classical strands. The cross-loading of MSA v. Dialect is also conceivable, since an ability to distinguish one from the other requires a familiarity with both. At the same time, the intercorrelations of the items within timepoints suggest that proficiency in these varieties is related to varying degrees, and not, as some suggest, distinct and entirely separable.

The fact that SR Proficiency scores correlate significantly with scores from the proficiency tests offers further evidence in support of the validity of the scales. The low coefficients between test scores and Dialect can be explained by the fact that the tests employed MSA only, which in comparison to Classical Arabic is further removed in terms of linguistic overlap. The higher coefficients for Y/N scores could be explained by the higher rate of transferability of vocabulary, in comparison to grammar, between varieties.

All of this supports the view that, whilst Arabic is an amalgam of multiglossic branches,

proficiency profiles differ when it comes to each variety, and attention should be paid to this heterogeneity of skills to avoid misrepresenting learners' abilities by solely presenting conflated proficiency scores.



4.2. The Nature of Motivational Self-Concepts

Research Question 2:

What is the nature of Arabic learners' motivational self-concepts? How do learners' motivational self-concepts vary based on:

- a. Arabic language variety;*
- b. Gender;*
- c. Heritage Learner status?*

This question was answered using both the quantitative survey data and the qualitative open-ended survey items and interviews. In keeping with the mixed-methods approach employed in this study, findings from each dataset were first analysed individually to provide responses to the questions above. The data were then interwoven using the 'Following the Thread' method (Moran-Ellis et al., 2006), where answers to questions raised by one dataset are sought in other sources.

In the sections that follow, findings from the quantitative data are presented first. This is followed by analyses of the interview data. Throughout each, 'Following the Thread' sections are included, where exploration which spanned and incorporated both datasets is detailed. These threads were identified through questions raised by one dataset for which there were interesting themes or through potential answers apparent in the other. Throughout the presentation of each dataset, corroborations and contradictions between the respective sets of findings are also detailed.

4.2.1. Quantitative Findings

4.2.1.1. Preliminary Analyses

4.2.1.1.1. *Completeness of Data*

A total of 270 students participated across the study. At T1, 265 took part from 19 different classes drawn from 16 schools. Of these participants, 173 (65.3%) were retained at T2, with an additional 5 participants joining this phase.¹⁸ The attrition was primarily due to 6 classes (from 4 schools, n=55) declining to continue their participation at a later point during the school year given the proximity of forthcoming end-of-year exams for Year 10 students. The remaining students who did not participate at T2 were students who were either absent from participating classes at T2 or simply individually declined to continue with the study.

4.2.1.1.2. *Attrition Bias*

Attrition bias, 'a type of selection bias due to systematic differences between study groups in the number and the way participants are lost from a study' (Nunan et al., 2018, p.21), can undermine the reliability of a study's findings. To determine whether there were any notable areas of difference between the retained sample at T2 and the 92 students who were no longer participating (the attrition group), descriptive statistics for both sets of learners were compared.

Of the 4 schools which did not participate at T2, only one was a boy's school (n=9), the remainder were all single-sex girls' schools. Of these, two were independent Muslim schools, one was a Muslim free school (containing two classes), and the last one was a non-faith academy (n=3) in which all students studying were from Muslim backgrounds. There does not appear to be a pattern amongst the schools which chose not to continue, and the gender distributions of classrooms largely reflect the original make-up of the sample: 15.8% of the classes were boys at T1, and 16.7% of the attrition

¹⁸ Of this total of 178 at T2, 175 completed the surveys, and 173 took proficiency tests. Neither figure totals 178, since at this time point, some students took only surveys and some only proficiency tests. The total for each set of individual analyses is provided in the output summaries in the Findings and/or Appendices. 175 is used to refer to the sample summaries since this was the number of students who completed the surveys at T2, which constitute the primary focus of analysis in this study.

classes were boys. Since schools in the sample were preparing students for the GCSE, content is likely to have been the same, and the type of school is unlikely to have affected the provision of instruction to a significant degree.

From Figure 4.5 it is evident that the two groups of participants (the retained sample and attrition group) do not differ substantially in terms of either their gender or heritage backgrounds, nor in terms of the amount of previous Arabic study. An independent samples t-test was also run to compare the average-score means for proficiency test scores and self-concepts for the two groups. No significant differences emerged indicating that there were no systematic differences between those who continued with and those who left the study. Thus, though the attrition sample was relatively large, the characteristics of the attrition group do not seem to differ from the original sample, and thus should not have skewed the make-up of the sample at T2 to such a degree that they may affect the findings.

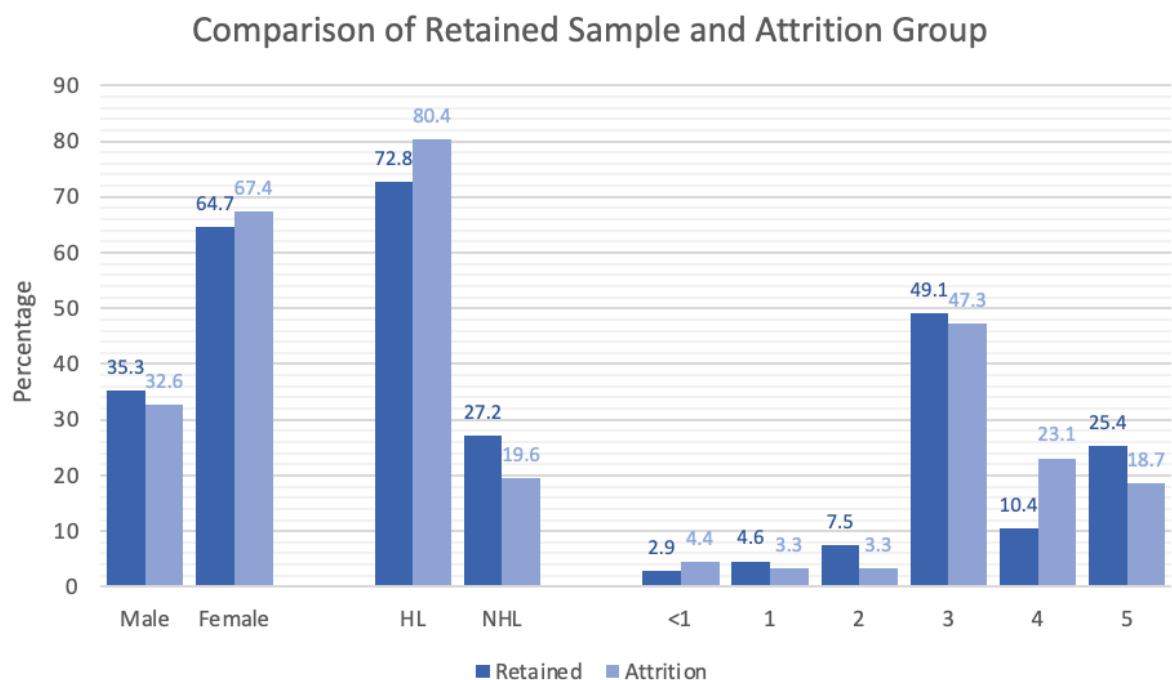


Figure 4.5. Bar graph comparing descriptive data of retained sample and attribution group

4.2.1.1.3. *Calculation of Average Scores*

In order to run mean comparison analyses, average-scores were calculated for all participants for each self-concept by combining scores for items pertaining to each latent variable (CS, IS, and OTS) and dividing by the number of items. Factor analyses had indicated the cross-loading of the Plausibility item (*'This is very likely—This is very unlikely'*) for each variety onto both the CS and IS. As a result, this was not included in the calculation of average-scores. Factor scores could not be used to run mean comparisons between variables and between sub-samples within the data since the scores for each are standardized around a mean of 0. Thus, to examine differences in the means between Arabic varieties and sub-groups, as well as across timepoints, these average-scores were used. The descriptive data for each of these average-scores is provided in Appendix C.15.

4.2.1.1.4. *Normality and Outliers*

In order to determine the nature of the data and subsequently the appropriate method of analysis, the data was examined for normality and outliers. Both z-scores and absolute scores of the skewness and kurtosis of the data were inspected. The acceptable range for z-scores for both skewness and kurtosis is $z > |3.29|$ for a sample of this size (Field, 2018). None of the z-scores for kurtosis fell outside of this range. Of the 18 variables however, five z-scores lay outside these limits for skewness at T1: *Dialect IS*, *Dialect OTS*, and *MSA OTS*; and at T2 *Dialect OTS*, and *MSA OTS*.

Since z-scores are not the only measure by which suitability for parametric testing can be ascertained, the absolute scores for skewness and kurtosis were also examined. According to Kline (2015), acceptable scores for skewness are <3.0 , whereas kurtosis scores should be <10.0 . It is important to note that scores within these limits do not qualify the data to be 'normally distributed' as true normal should be 0 for both; however, these qualify as not 'severely' non-normally distributed and therefore suitable for parametric tests. Since, according to Kline's criteria, the data was still suitable for parametric testing, this was employed.

4.2.1.2. Comparison of Self-Concept Average-Scores

4.2.1.2.1.1. Whole Sample

Within-student differences in the three self-types (Ideal, Current, and Ought to) across the three language varieties (Classic, Modern Standard, Dialect) were first investigated using Bonferroni corrected paired t-tests ($p < .006$). The same pattern—*Ideal > Current > Ought-to*—was reported in all three language varieties.

At each timepoint, the average-score means for each self-concept component were also compared between language varieties using the paired t-tests. Results indicated that although the same self-concept hierarchy emerged within each language variety, scores for individual self-concepts differ significantly from one language variety to another (See Appendix C.16 for full results). Figure 4.6 presents these scores, arranged by self-concept. Each box contains the variety mean for that self-concept. Bold arrows between boxes (provided along with the correlation coefficients) represent significant differences between language variety means.

Current Self

T1: Classical > Dialect > MSA

T2: Classical = Dialect > MSA

For Current Self, the scores for each variety were significantly different from one another at T1 (see Figure 4.6). The mean score for MSA (3.00) was significantly lower than the 3.34 mean score for Dialect ($t(259) = 6.39, p < .005, d = .40$), which in turn was lower than Classical, which had the highest mean of 3.52 ($t(259) = 6.39, p < .005, d = .21$). This is likely due to students' familiarity with reading religious texts. The Cohen's d effect sizes for both of these differences were small according to the benchmarks suggested by Cohen (1988). The effect size for the difference between MSA and Classical was medium ($d = .56$). It is worth remembering that these means are from a total of 6 Likert-scale points and thus, although learners' Classical Current Selves emerged as highest, they are still only slightly above the halfway point on the scale.

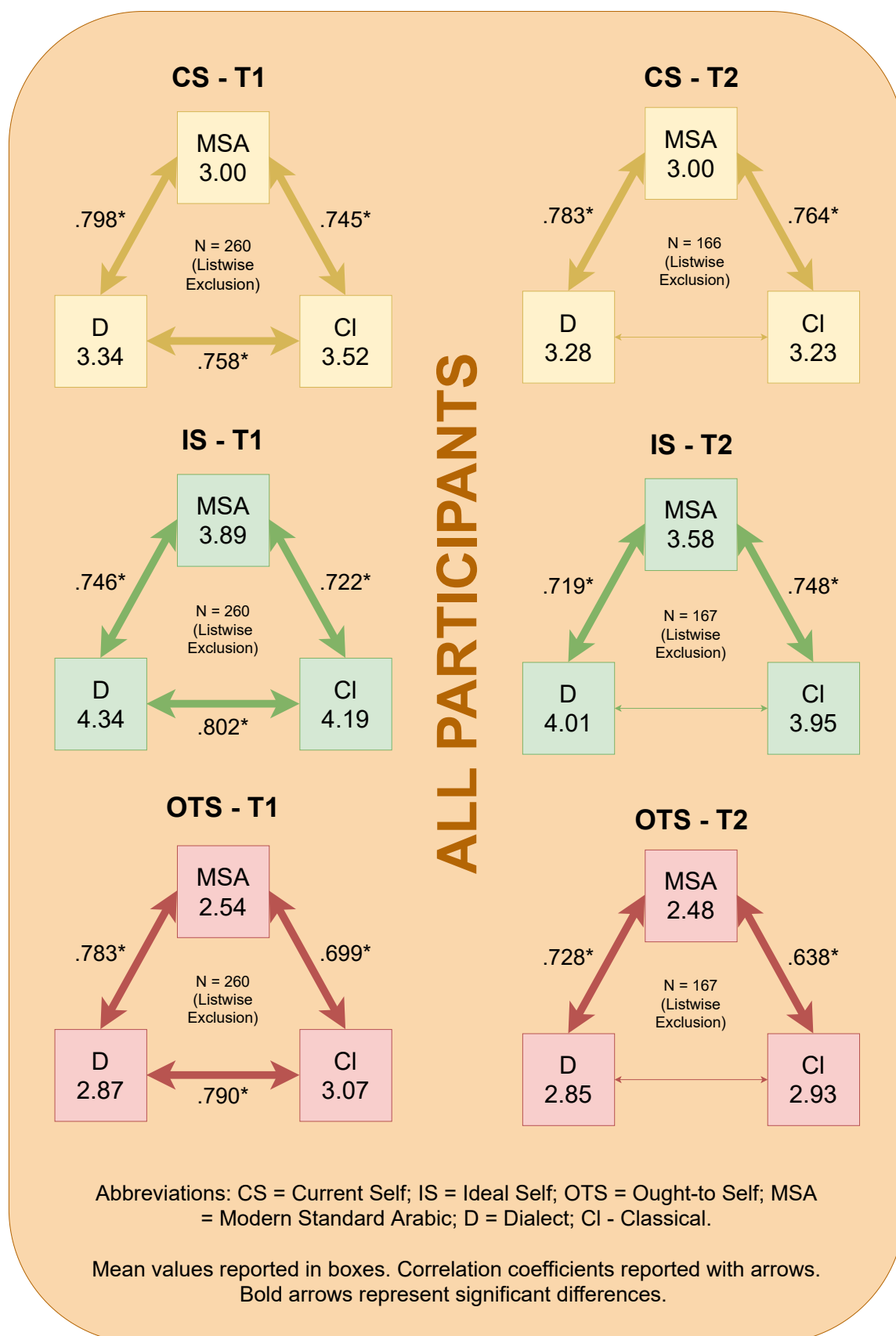


Figure 4.6. Comparison of mean scores per self-concept for each variety for the full sample

At T2, MSA CS scores also emerged as significantly lower than both the Classical ($t(165)=-3.12, p<.006, d=.24$) and Dialect varieties ($t(166)=3.75, p<.0005, d=.28$); however there does not appear to be a significant difference between the latter two scores. Effect sizes for both these significant differences were small. The lack of a significant difference between Dialect and Classical CS scores at T2 appears to have been caused by a fall in both to within closer range of each other, indicating a weakening in both self-concepts across time. Appendix C.16 contains the complete t-test statistics.

Ideal Self

T1: Dialect>Classical>MSA

T2: Dialect=Classical>MSA

For IS at T1, a slightly different hierarchy emerges, with scores for Dialect being the highest, followed by Classical, and finally MSA scores being the lowest, with each of these being significantly different from one another. At T2, a similar fall in the average-scores for Dialect and Classical IS to within closer range of each other appears to have precipitated the same pattern as that of the CS. The mean average-score for MSA IS remains significantly lower than both Dialect ($t(166)=4.66, p<.0005, d=.36$) and Classical Arabic ($t(166)=-4.24, p<.0005, d=.33$), whilst the difference between the means for each of these is no longer statistically significant (see Figure 4.6). Again, all Cohen's d effect sizes for these outcomes were small. See Appendix C.16 for full t-test results and effect sizes.

Worth noting is the fact the means for each of these varieties were higher for IS (ranging between 3.58 and 4.34) than for CS (which ranged from 3.00 to 3.53) at both timepoints, which is to be expected since, according to the L2MSS, motivation is derived from the desire to reduce the discrepancy between the CS and the IS.

Ought-to Self

T1: Dialect>Classical>MSA

T2: Dialect=Classical>MSA

In contrast to the other two self-concepts, the scores for OTS were consistently lower than for the other two self-concepts, with the means ranging from 2.48 to 3.07 across timepoints (see Figure 4.6). Nevertheless, the same pattern as IS emerges with respect to the hierarchy of varieties: at T1 the mean for the MSA OTS is significantly lower than that of Classical ($t(259)=-9.47, p<.0005, d=.59$). The effect size for this difference was medium. Classical was in turn significantly lower than the mean score for Dialect ($t(262)=4.25, p<.0005, d=.26$).

Similarly, the mean scores for each variety fell at T2, indicating a weakening of the OTS concepts over time amongst participants. The lack of significance in the difference between Dialect and Classical Arabic appears to have been caused predominantly by the change in the mean for Classical which fell from 3.07 at T1 to 2.93 at T2, in comparison to Dialect which changed by only -.02 across the timepoints. The effects sizes for both these differences were in the small range. See Appendix C.16 for full output.

4.2.1.2.2. Summary

What is evident from these cross-variety comparisons of the sample as a whole is that the IS scores are consistently higher than CS, which are in turn higher than those for the OTS, for each language variety, as depicted in Figure 4.7. In addition, it is apparent that scores for MSA for all self-concepts were significantly lower than both Dialect and/or Classical Arabic.

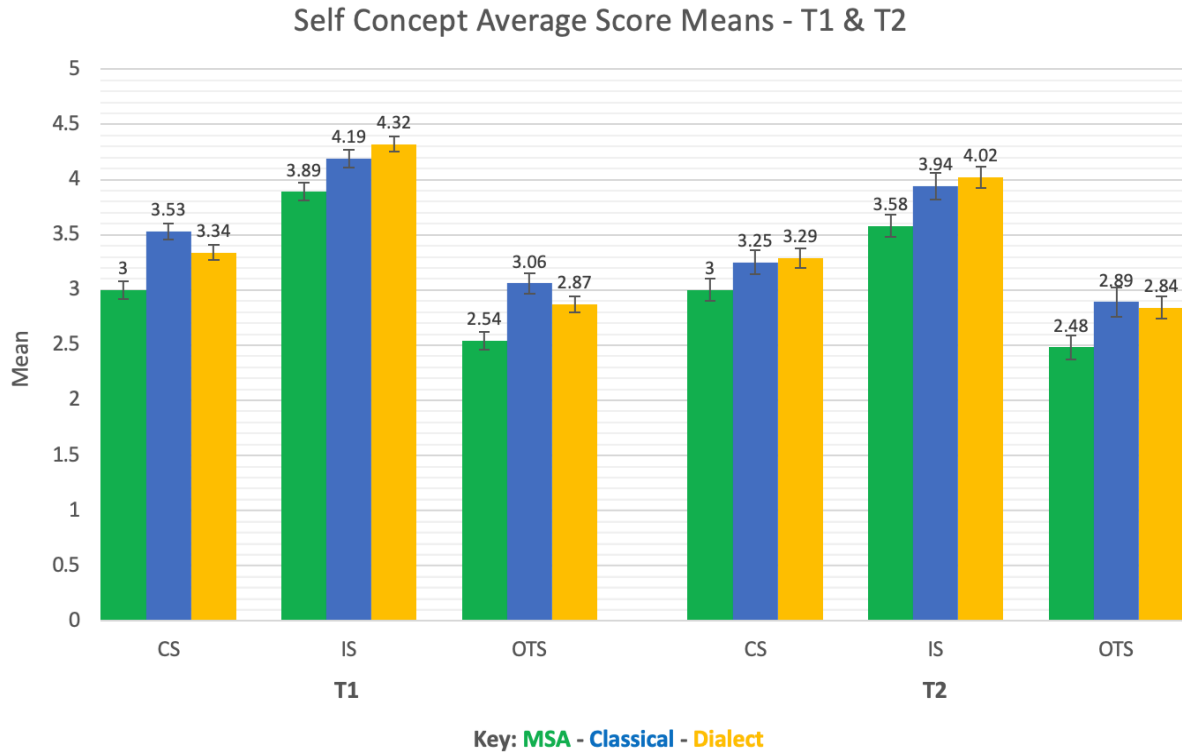


Figure 4.7. Bar graph comparing self-concept average-score means per variety at T1 & T2.¹⁹

4.2.1.3. Following the Thread: Analysing the Variety Hierarchy

In order to offer an explanation for the emergent pattern of statistical findings, the interview data was explored for specific references to the varieties as described in the surveys:

- MSA: *fully understanding the news, books, or magazines.*
- Classical: *fully understanding religious texts in Arabic.*
- Dialect: *communicating fluently with Arabic speakers.*

All references to these specific uses of Arabic were identified and then sub-coded into references to CS, IS, OTS, and other categories (such as Past Self, see Section 4.2.9).

¹⁹ 95% Confidence Intervals for each variable are represented by error bars included in the graph. This is repeated for each bar graph of variable mean scores.

From the frequencies in Table 4.6, it is clear that explicit mentions of the specific uses of MSA by students were far fewer for all self-concepts than the other varieties. This sparseness of references may reflect an apparent lack of interest in MSA (or perhaps a lack of awareness of its application) which may explain why self-concepts scores for this variety were significantly lower than the others.

Table 4.6. Frequency of Cases and References to Varieties		
Variety	Cases Coded	Coding References
Classical	67	295
Dialect	64	317
MSA	35	95

Table 4.7 provides a cross-tabulation of references to variety-specific mentions of self-concepts. The total number of OTS references related to the specific varieties was 48. The frequency of all OTS references across the dataset was 320. This indicates that the vast majority of OTS references were made in relation to non-specific language goals, or to the GCSE course, rather than any of the three language varieties.

Table 4.7. Cross-tabulation of variety-self references				
Variety	Classical	Dialect	MSA	Total
Current Self	58	39	13	110
Ideal Self	78	119	17	214
Ought-to Self	38	9	1	48
Total	174	167	31	372

The low scores for MSA, which is the variety learners are studying for their GCSE is, on the surface, surprising. It appears, as a result, that the GCSE does not map onto the specific applications of pure MSA in authentic settings in the minds of learners. In addition, it can be inferred that the actual application of pure MSA is neither something that appeals to learners, hence the lower IS scores, nor, in fact, do learners feel any obligation to achieve the MSA-related future self-concept, as expressed by the almost complete absence of OTS references in this regard.

4.2.1.4. Following the Thread Further: Feelings Towards Arabic v. GCSE

4.2.1.4.1. *T-Tests*

The distinction between the GCSE subject and the Arabic language itself was included in the survey. Learners were asked to rate their feelings towards the Arabic language and towards the GCSE on two 6-point Likert-scales.

To determine whether there was indeed a difference in how learners felt towards the GCSE, as opposed to the language, a paired-samples t-test was run on both the T1 and T2 responses to these items. At both timepoints, learners' scores for their feelings towards the language ranked higher than those towards the GCSE. Further, at both timepoints participants' responses to each item were statistically different, as illustrated in Table 4.8, though the effect sizes were in the small range.

In tandem with the frequencies from the interview data in Tables 4.6 and 4.7, these scores suggest a disjuncture between learners' conceptions of the Arabic language as a whole, which incorporates multiple varieties and applications, and the more restricted, oftentimes enforced, GCSE subject.

Table 4.8. Paired Samples T-Test results—Comparison of Feelings towards Arabic & GCSE

		Mean	N	Std. Deviation	Std. Error Mean	Mean	Std. Deviation	Std. Error Mean	95% CI		t	df	Sig. (2-tailed)	Effect sizes (Cohen's d)
Pair 1	Feelings towards Arabic T1	4.68	263	1.416	0.087	0.395	1.024	0.063	0.271	0.52	6.261	262	.000***	.38
	Feelings towards the GCSE T1	4.29	263	1.472	0.091									
Pair 2	Feelings towards Arabic T2	4.51	173	1.512	0.115	0.497	1.082	0.082	0.335	0.659	6.045	172	.000***	.46
	Feelings towards the GCSE T2	4.01	173	1.607	0.122									

A further potential reason for the difference between the varieties in terms of the individual self-concepts is the lack of experience amongst participants in applying skills to the aims of the varieties as stated in the survey. The application of MSA was expressed in the opening statement as *'fully understanding the news, books, or magazines'*. Only once did a student mention that they had attempted to read materials in MSA not linked to their exams:

'Even now I went on Emirates, and you know, they read the instructions out in Arabic, and I understood some of that, and I was telling my mum.'

Though this is not a specific reference to news, books, or magazines, it does make mention of authentic reading material. Blair and Azaz's (2019) recent research suggests there are links between past selves and ideal selves. Memories of real-life experiences can lend themselves to promoting an IS concept. Hence, the lack of references to experience with reading authentic material may offer a potential explanation for the low MSA self-concepts frequencies and the finding that it is significantly lower than the other two varieties in each domain.

A number of other noteworthy findings emerged from an analysis of these self-concept codes for each variety in the interviews, which further corroborate findings from the quantitative data. The first relates to the relationship between the CS and IS. The existence of learners' current abilities appears to lend itself to promoting their Ideal Selves. For example, in the context of Classical Arabic, learners expressed: *'you read it all the time and you want to be able to understand it [the Quran]'*; and *'when I read the Quran, I really want to understand the meaning of it'*. The pairing of the present tense with the use of the modal phrasing 'want to be able to' indicates a continuum or overlap between the two, with the CS driving the IS. Similarly, for Dialect: *'I think I'm okay with like, conversations and stuff... But obviously I still need to learn more. Um, but yeah. I do see myself using it, I do see myself, like you know, progressing in Arabic.'* In this case, the CS might be conceptualised as part of the reason for the existence of the IS. In other words, because learners can already utilise some of their language knowledge, they may be keener, or more motivated, to develop this into a fully-fledged future self. In essence,

in the case of MSA, learners do not express high levels of CS concepts, which in turn could then explain the low IS scores.

Another way to understand the relationship between CS and IS is through the lens of *plausibility*. This was originally conceptualised by Dörnyei (2009) as a component of the IS. However, this variable cross-loaded onto both the CS and IS in the Factor Analysis in Section 4.1.1.4.1. The relationships observed in the interview data in this section appear to suggest that plausibility consists of current realities as a result of which a future self is seen as achievable. In other words, plausibility is constituted of current possibilities and the related development of these existent possibilities, hence the cross-loading. Whilst these patterns are not observed for all learners, they do, at the very least, highlight the necessity of examining the interaction between motivational self-concepts, and not just the discrepancies between them.

A further finding, which again corroborates the aforementioned results relating to the OTS, particularly in regards to Classical Arabic, is the overlap between IS and OTS. This is indicated by the use of words such as ‘necessary’ and ‘better’ in learners’ answers, made apparent by examination of the variety specific self-concepts.

Student: ‘It’s more better [sic], I feel, if you read it [the Quran] in Arabic.’

Researcher: ‘Why is it better?’

Student: ‘Because it was revealed in Arabic. So when you read it in Arabic, you’re reading it the way it was revealed.’

An indication of this overlap is the use of the word ‘need’: ‘To read the Quran, you need to know Arabic.’ This is not an explicit statement of obligation; however, it does suggest an importance attached to Arabic, which participants recognise and value. This overlap is alluded to, in a less direct sense, through references to the reasons participants give for wanting to understand Quranic Arabic:

... you just had to know how to read it [the Quran]. But I think, um, when-- obviously, what you should try to do to understand the Quran better is to, um, learn

the words. So that's why I wanted to-- wanted to learn the actual language and what it means...

The use of 'should' and 'want' in succession indicates that an underlying sense of obligation can lend itself to the development of an IS. This makes sense; religion is often a source for a person's ideals. It does, however, raise questions about the possibility, and even necessity, of identifying a clear distinction between these two future self-concepts.

4.2.1.5. Summary

In sum, the findings for the sample as a whole indicate that the anticipated pattern of IS being higher than CS does indeed emerge at both timepoints for all three language varieties. Unexpectedly, the OTS scores are significantly lower than both. In addition, scores for MSA for each self-concept emerge as significantly lower than both of the other two varieties. This is unexpected since this is the variety learners are studying for their GCSE. A possible explanation for this lies in the actual authentic uses to which MSA is applied—reading modern formal written texts—as outlined in the survey scale opening statement for this variety. The dearth of references to this context of use in the interviews offers an explanation as to why learners (according to their low IS and OTS scores) neither desire nor feel obliged to achieve this future self-concept; they are either uninterested or unaware of how MSA is utilised. A comparison of learners' self-reported feelings towards Arabic and towards the GCSE reveals significant differences between the two, indicating that participants view these two constructs (the language and the exam) differently. The themes raised by the Following the Thread investigations also highlight the issue of the interrelatedness and potential overlap of the self-concepts, indicating the necessity of employing both caution and paying closer attention not just to the distinctions between the constructs but also the interplay between them.

4.2.2. Gender Differences

4.2.2.1. Descriptive Analyses

At both T1 and T2, the proportion of males in the sample ($n_{T1}=85$; $n_{T2}=58$) was smaller than females ($n_{T1}=179$; $n_{T2}=116$). Since both sub-sample sizes were within the 50-300 range, $z>|3.29|$ remained the acceptable cut off points for skewness and kurtosis to qualify for parametric testing. The z-scores for all factors for males at T1 and T2 were within this range for both skewness and kurtosis. Whilst the kurtosis z-scores for females were all within this limit at T1 and T2, three items exhibited skewness z-scores outside of this range (*Dialect OTS T1*; *MSA IS T1*; *MSA OTS T2*). An examination of the raw scores for all these items, however, show them to be well within the acceptable limits for parametric testing (Kline, 2015), which was thus employed. See Appendix C.17 for the full set of descriptive statistics.

4.2.2.1.1. *Between Genders*

The mean scores indicate that, for all varieties and self-concepts, males scored higher than females. This pattern was repeated at T2. To determine whether these raw mean differences were significant, a set of Bonferroni-corrected independent samples t-tests were run, the adjusted minimum p -value for which was .006. None of the mean differences between genders emerged as significant, thus providing no evidence in this data that gender has an effect on learners' self-concepts in this context (see Appendix C.18 for full t-test results).

4.2.2.1.2. *Within Genders*

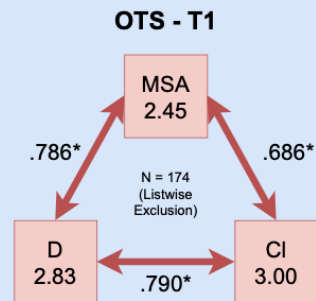
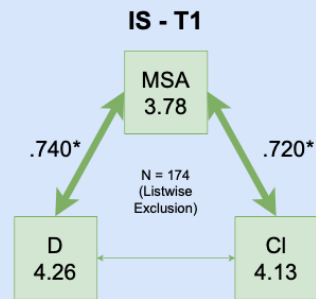
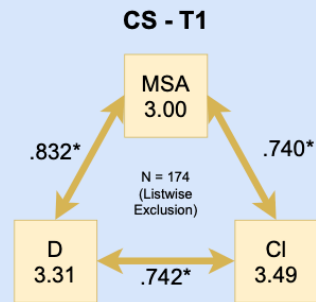
In order to establish whether there were significant differences between the varieties within the scores within each gender, the data file was split first by gender, and then paired samples t-tests were run on these separated data sets. A Bonferroni adjusted p -value ($<.008$, i.e., $.05/6$ tests) was employed for these tests, since six t-tests were run together at the same time. According to the results, the means for MSA were consistently significantly lower than the other two varieties across each domain of self-concept for both males and females at T1. The exception to this pattern was the IS for

MSA for males, which was significantly lower than Dialect ($M=0.35$, 95% CI [0.14, 0.57], $t(81)=3.329$, $p<.001$, $d=.36$), but was not lower than Classical. This is summarized in Table 4.9.

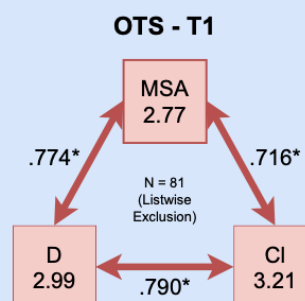
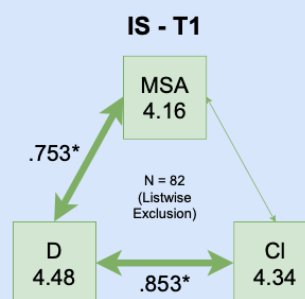
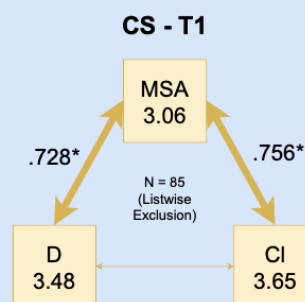
Table 4.9. Language Variety Hierarchies in Male and Female Subgroups			
	Self-Concept	Males	Females
T1	Current Self	Classical=Dialect>MSA	Classical>Dialect>MSA
	Ideal Self	Dialect>Classical=MSA	Dialect=Classical>MSA
	Ought-to Self	Classical>Dialect>MSA	Classical>Dialect>MSA
T2	Current Self	Dialect>MSA Dialect=Classical MSA=Classical	Classical=Dialect=MSA
	Ideal Self	Dialect>MSA Dialect=Classical MSA=Classical	Dialect=Classical>MSA
	Ought-to Self	Classical=Dialect>MSA	Classical=Dialect>MSA

This pattern diverges somewhat at T2 with MSA significantly lower than Classical and/or Dialect for IS and OTS. However, for CS, the only significant difference is between Dialect ($M=3.66$, $SD=1.37$) and MSA ($M=3.09$, $SD=1.47$) with the scores for Dialect being significantly higher ($M=0.55$, 95% CI [0.29, 0.80], $t(54)=4.267$, $p<.001$, $d=.58$) amongst males students. No evidence of significant differences were apparent between the varieties for female students. The t-test results are represented visually in Figure 4.8 (see Appendix C.19 for full output). Arrows representing significant differences between varieties are depicted in bold, accompanied by the correlation coefficients. The majority of the effect sizes calculated for each of these significant differences were small, with only 6 of the 23 variables for which significance emerged falling within the medium effect size range.

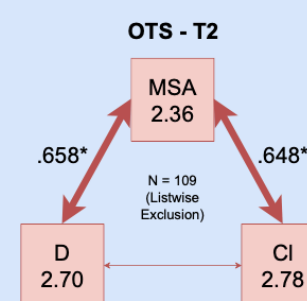
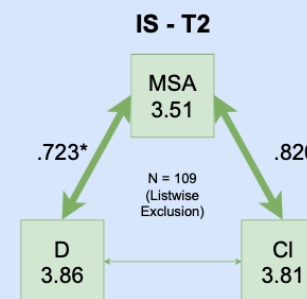
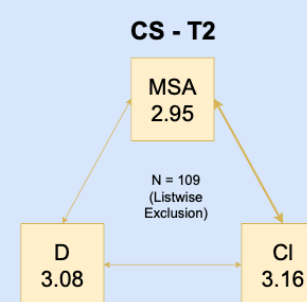
Females



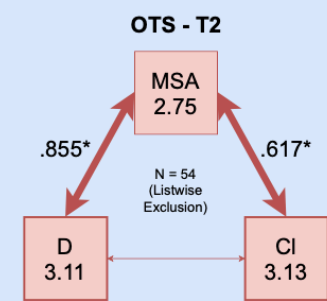
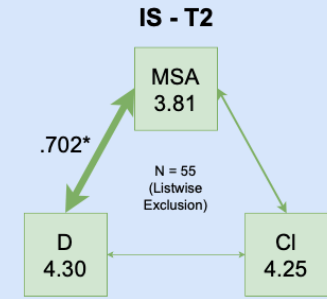
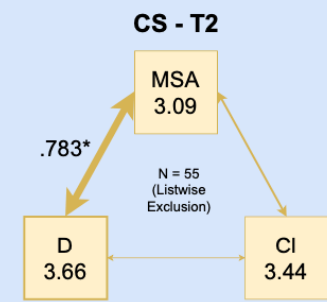
Males



Females



Males



Abbreviations: CS = Current Self; IS = Ideal Self; OTS = Ought-to Self; MSA = Modern Standard Arabic; D = Dialect; CI - Classical.

Mean values reported in boxes. Correlation coefficients reported with arrows. Bold arrows represent significant differences.

Figure 4.8. Comparison of mean scores per self-concept for each variety for males and females

4.2.3. Heritage Background Differences

4.2.3.1. Descriptive Analyses

The non-heritage learner (NHL) proportion of the sample was approximately three times the size of the heritage learner (HL) proportion at both T1 and T2. Given the size of the HL sub-sample at T2 ($n=42$), a lower cut-off z-score limit of $z>|1.96|$ was employed in line with Field's suggestions (2018). At T1, the z-scores for only one factor lay outside the acceptable limit (MSA OTS). At T2, none of the NHL z-scores suggested non-normality, whereas for HLs three average-score z-scores exceeded the new stricter limit (*Classical IS*; *Dialect OTS*; and *MSA IS*). However, all raw scores were well within the acceptable limits for parametric tests according to the limits suggested by Kline (2015). See Appendix C.20 for full descriptive statistics.

4.2.3.1.1. *Between Backgrounds*

The raw mean scores for Current Selves across all varieties are higher for HLs, which is expected given their experience and affiliations with the language (Olsen, 2018). A striking finding, however, is that the means for OTS, however, are higher for NHLs at T1. At T2, HL mean scores are higher than all NHL scores across self-concepts and varieties.

Independent samples t-tests however (adopting the adjusted p -value of .006) indicate that at T1, there were no significant differences between these two sub-groups. Results are illustrated in Figure 4.9 and listed in full in Appendix C.21. This suggests that HL status at the beginning of Year 10 has no significant impact on learners' self-concepts concerning any language variety. At the end of the academic year, however, this has changed substantially. For five out of the nine factors, the difference between HLs and NHLs is now significant at $p<.006$. For the following four variables, the differences between the average-score means were not significant: *Classical OTS*; *Dialect IS*; *Dialect OTS*; *MSA OTS*.

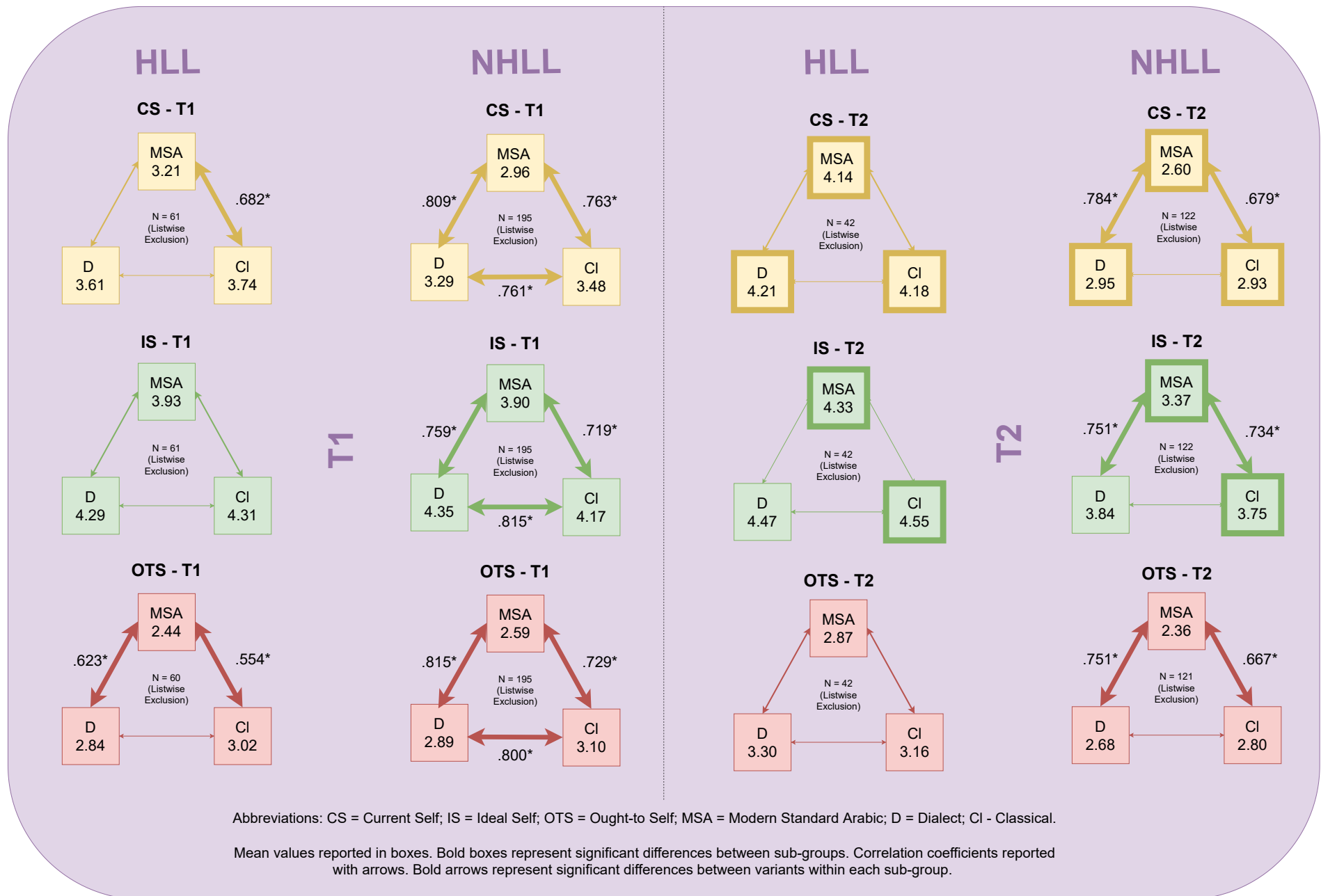


Figure 4.9. Comparison of mean scores per self-concept for each variety for HLs and NHLs

For all other average-scores, HLs exhibited a significantly higher mean score. The largest mean difference was for MSA CS, with HLs scoring much higher on the Likert-scale ($M=4.143$, $SD=1.178$) than NHLs ($M=2.604$, $SD=1.178$), a mean difference of 1.539 (95% CI [-1.98, -1.10], $t(162)=-6.887$, $p<.006$). These scores suggest that, though HLs in this study did not have stronger self-concepts than NHLs at the beginning of their first GCSE year, at T2, many dimensions of their self-concepts were significantly stronger across all three varieties.

Both groups appear to feel as equally obliged at T2 to learn all three varieties. In the realms of IS, both groups feel equally keen to speak the Dialect variety; background appears to have no significant impact on this desire. Across the varieties however, HLs at T2 appear to feel significantly more confident and competent than NHLs, since the means for the CS concepts are invariably higher.

Accordingly, one may argue that learners' heritage backgrounds play a significant role in the nature and development of their motivational profiles. At the same time, it is clear that learners' profiles in this dataset are distinctive for different varieties indicating the danger in combining these into an Arabic monolith when discussing motivation.

4.2.3.1.2. *Within Backgrounds*

The sample was split on the basis of HL background, and raw scores were once again examined. A similar pattern to that exhibited for the whole sample was present in each group, with the mean MSA score constituting the lowest of the three varieties for both groups at T1.

Paired samples t-tests were run to establish whether, within each group, there were significant differences between the varieties for each domain of self-concept. As before, an adjusted significance level of .008 was set using a Bonferroni calculation. Results are depicted in Figure 4.9. Of the total pairs, 19 emerged as significant, the effect sizes for which ranged from .27 to .63 (see Appendix C.22). The hierarchies that emerged are outlined in Table 4.10:

Table 4.10. Language Variety Hierarchies in HL and NHL Subgroups			
	Self-Concept	HLs	NHLs
T1	Current Self	Classical=Dialect Dialect=MSA Classical>MSA	Classical>Dialect>MSA
	Ideal Self	Classical=Dialect=MSA	Dialect>Classical>MSA
	Ought-to Self	Classical=Dialect>MSA	Classical>Dialect>MSA
T2	Current Self	Dialect=MSA=Classical	Dialect=Classical>MSA
	Ideal Self	Classical=Dialect=MSA	Dialect=Classical>MSA
	Ought-to Self	Dialect=Classical=MSA	Classical=Dialect>MSA

The T-test results at T1 offer no evidence for the existence of significant differences between the OTS scores for Classical and Dialect Arabic for HLs. Neither does the difference between Classical and Dialect IS for HLs emerge as significant.

For all self-concepts, MSA is significantly lower than both other varieties in both groups except for HLs' IS concepts for which there are no significant differences between the varieties, and for their CS concepts, for which the MSA mean was significantly lower only than Classical.

At T2, different patterns emerge within each group. For NHLs' CS concepts, MSA is significantly lower than both Dialect ($M=0.33$, 95% CI $[-0.30, 0.45]$, $t(123)=4.702$, $p<.001$, $d=.42$) and Classical ($M=-0.32$, 95% CI $[-0.49, -0.15]$, $t(122)=-3.778$, $p<.001$, $d=-.34$). For HLs on the other hand, there was no evidence that the scores for each variety were significantly different.

For the IS, again, none of the differences between the variety means emerged as significant for HLs. Within the NHL group, MSA is significantly lower than the other two varieties. The same pattern is evident for OTS. In other words, HLs in this dataset appear to feel equally obliged and keen to realise future states pertaining to all three varieties of Arabic. NHLs in this sample feel keener and more obliged to be able to communicate with native speakers and to understand religious texts than they do to be

able to access modern written material. This is perhaps explained by the awareness amongst HLs of the necessity of MSA in real life settings, whereas NHLs, given their lack of exposure to authentic uses of Arabic, do not consider the use of this variety to be a desirable nor obligatory component of their future selves.²⁰

The differences between varieties across the self-concepts demonstrates once again the necessity of treating them as distinct language phenomena, rather than conflating them under the umbrella of Arabic as a whole. Furthermore, it offers evidence for differences in the motivational profiles of HL and NHLs within GCSE classrooms.

Most striking, perhaps, is the fairly consistent emergence of significantly lower scores for MSA across the motivational self-concepts for both groups (although the effect sizes are small-medium). This variety is generally used in formal written texts, hence the opening statement in the survey pertaining to reading the news, books, and magazines. Since the GCSE is focused on the Modern Standard variety, these findings have important implications for curriculum design and pedagogy. Learners' IS and OTS concepts are clearly stronger for the colloquial and classical varieties, neither of which are incorporated into the GCSE syllabus.

4.2.4. Qualitative Findings

The combined *a priori* and exploratory coding approach, outlined in the Methodology, was employed in the analysis of the interview data and open-ended survey items. In total, 95 students participated in the group interviews. The findings presented in this section provide an in-depth exploration of the motivational self-concepts as interpreted by the participants, accompanied by illustrative quotations from the interviews.²¹ These findings corroborate, as well as challenge in areas, findings from the quantitative data, in addition to Dörnyei's (2009) original, underlying conceptualisation of motivational

²⁰ When it comes to MSA material, there is also the possibility that any differences between heritage groups may relate to learners' identities and interests. It may be that HLs from Arabic-speaking backgrounds consider Arabic news material to be of interest culturally, and indeed trustworthy sources of news, whereas NHLs from other ethnic backgrounds may not see these sources as interesting or relevant. The validity of this explanation requires further investigation in future studies.

²¹ Note that some of the quotations are coded in multiple categories and so may recur in different sections of the analysis.

selves. References to each self-concept are examined and presented in turn, and relevant emergent differences between genders and heritage/non-heritage learners are outlined.

4.2.5. 'Different Arabics'

Before embarking upon an analysis of the motivational self-concepts in the interview data, attention was first turned to examining the validity of the tripartite taxonomy of the Arabic language utilised in the survey data. Research into the linguistic nature of Arabic has revealed multiple varieties of the language, all of which are referred to interchangeably as 'Arabic', and knowledge of all of which constitute, to varying degrees, fluency in the language. The tripartite delineation of Arabic used in this study identified Modern Standard, Classical, and Colloquial Arabic as representing the three major varieties. As discussed in the Literature Review, these are, however, less discrete categories than poles of a multifaceted spectrum, along which an individual will locate the form of language they are employing depending on the specific context of use.

The ecological validity of this taxonomy was indeed corroborated by learners' own references to 'different Arabics'. Whilst it could be argued that exposure to the survey instruments facilitated this view of the language, learners' responses clearly included their own pre-established terminology to refer to different language varieties, as will be explored in the following sections.

4.2.5.1. 'Worldly' Arabic (MSA)

Learners made the distinction between Quranic Arabic (described below) and the Arabic learned in school for the purposes of the GCSE, which they referred to with different phrases: 'worldly', 'everyday', and 'general'. Nine references used such terminology to explicitly describe the latter type of Arabic (see Appendix C.23 for full list of references). It appeared that the perception of this difference was based largely on vocabulary and topic coverage, rather than grammatical differences. One learner made the comment that, in her attempts to apply her GCSE Arabic knowledge to Quran comprehension, she did not see any overlap in vocabulary items. Ironically, the item she

chose as an example اسمي, *ismi*, the root of which means ‘name’ is used at the start of each of the 114 chapters of the Quran (bar one) albeit in a slightly modified form, in the phrase ‘In the Name of God’, which Muslims also use in everyday life. Whilst this particular student’s statement was an example of her feelings of despondency and frustration with the GCSE, this perceived lack of transferability between GCSE and Quranic Arabic was not uncommon in the data. This raises issues both in terms of the utility of the exam content in relation to learners’ motivations, and the lack of focus on skill application within Arabic classrooms.

I want to study Arabic religiously... Quran and Hadith [sayings of the Prophet Muhammad], I want to understand those aspects of Arabic. We’re doing more of news and sports, and that type of stuff so I don’t think that’d really benefit me a lot in the future.

Examples of where skill application was present were highlighted favourably by students, providing evidence of its possibility and appreciation.

Student 1: Not all the words in the Quran are words that we learned. But there are still certain words-- even just some general words, like, the ‘earth’, you can still pick that out when you’re reading the Quran and sometimes I can understand... one āyah [verse] what it means, so yeah, I think that it really helps me.

Researcher: How does that make you feel, when you can pick out those words?

Student 1: It makes me feel happy, like--

Student 2: And accomplished.

4.2.5.2. ‘Mosque Arabic’ (Classical)

Numerous terms were used to describe Arabic used for religious purposes. These included: ‘religion-based’, ‘Islamic related’, ‘Quranic’, and even ‘ancient’ Arabic. Fourteen references were made describing this type of Arabic (see Appendix C.23).

References to this type of Arabic were commonly made in the context of discussions about the ‘Alimah course and language courses taken in the mosque. These courses invariably focused on the comprehension of religious texts and the use of Arabic in the

study of Islamic sciences. Comments on GCSE Arabic focused on the lack of lexical overlap with Classical Arabic. Contrastingly, in elaborations of their experiences with ‘mosque’ Arabic, participants spoke of their advantages in grammatical knowledge acquired in Classical learning contexts when it came to the GCSE. One student, for example, spoke of observing her classmates being taught, and struggling to learn, verbal conjugation patterns, which she herself had covered in the mosque. She contrasts the teaching methods—the traditional presentation and memorisation of the conjugation table (consisting of 14 conjugated forms) in the mosque with the, inferably, communicative method whereby these forms are learned less systematically—and expresses that the approach she enjoyed in the mosque was preferable:

The way it's structured, our mosque books... or example, when we're taught verbs... The way we learnt, there's 14 [pronouns and conjugated verb varieties], right? They get taught it in a different order—When they get taught... It's, like, they learn the masculine one. I don't know, they have a different order. If they were to use my table I feel like it'll be so much easier. Like, I see them struggling with verbs.

Those students studying Arabic in the mosque also spoke of enjoying that particular course more than the GCSE, and of doing better in these extracurricular studies. This was attributed to the personal relevance of content covered, the style of instruction, and in some cases to the belief that mosque Arabic had helped them with the GCSE. In the context of the L2MSS, this would represent a process of realising their IS and reducing the gap between learners' current and future self-concepts.

Whilst there was a level of grammatical overlap from Classical to MSA (taught in the GCSE classroom), the dearth of reverse transferability from GCSE Arabic to mosque Arabic, plus the difference in teaching methods, appears to add to the perception that the Arabic covered in the mosque was ‘completely different’ to that covered in school, a view corroborated in one case by the student's mosque instructor:

She said that even though she's an 'Alimah, she understands Arabic, she wouldn't teach it—Arabic GCSE. She says that Arabic—the Arabic is different to mosque Arabic, like, what we learn in mosque. Two different types of Arabic.

4.2.5.3. 'My Arabic' (Colloquial)

Least frequently mentioned was the colloquial variety of Arabic. References including explicit descriptions of this type of Arabic totalled seven (see Appendix C.23). Students who referred to this type of Arabic were mainly HLs, who made mention of the specific dialects they spoke at home with phrases such as 'Algerian dialect' (in contrast to 'normal Arabic') and 'my Arabic dialects'. However, some learners from NHL backgrounds also mentioned an awareness of different dialects (e.g., '*all the other Arab countries, they have different dialects, so it does get a bit confusing*') and the desire to learn different types of Arabic, including dialects, which one student referred to as 'slang', to be able to use them for different purposes:

... Just different ways of Arabic, like, colloquial, and different stuff, so I do wish that we learnt more different styles of Arabic. Not just one way. But ... even, like, in Arab countries, a lot of people talk in slang Arabic as well, different ways, so it'd be helpful if we learnt all types.

Overall, the fact that students within this sample identified different varieties of Arabic which overlap with these presented in the theoretical framework of this study adds validity to this delineation. The infrequent references to colloquial Arabic, however, suggest that learners are not fully aware of the linguistic diversity existent in settings of authentic Arabic usage. The furthest they are able to go is to express difficulty in understanding Arabic speakers, given their accents and speed of speech, or the recognition of differences between the Arabic used in their homes, in the case of HLs, and that learned in the classroom. The starkest difference participants perceive lies between MSA and Classical Arabic. In reality, the overlap between these two varieties is much larger than the overlap between MSA and Dialect, which differs grammatically as well as lexically. But in the context of these learners' intended usage, the apparent lack of lexical overlap, or at the very least their lack of experience in applying their knowledge of MSA to religious texts, exaggerates the differences between the two, and raises questions of the utility and efficacy of the GCSE course.

4.2.6. Current Self

The first self-concept to be explored was the Current Self. Learners provided details regarding their CS concepts, both in response to interview questions designed to specifically elicit this information and in their answers to questions concerning other components of their self-concepts. Learners were asked about their experiences with learning Arabic—a broad question designed to both set the context for the rest of the interview, and to help participants ‘warm up’ and become comfortable with the researcher before more specific questions were asked. Participants were next asked to speak about their current levels of confidence in Arabic. The wording of this particular question was left purposefully open to allow learners to respond with their perceptions of confidence as pertaining to the Arabic language itself, as well as with regards to the GCSE curriculum and their language lessons.

CS in this study was operationalised so as to incorporate learners’ *confidence* with regards to the language, as well as their current *self-assessed ability* in a particular domain of language use. These components were addressed by the interviewees to varying degrees of detail, which are discussed below. Participants’ discussions concerning their Current Selves in relation to Arabic, however, also incorporated their *perceptions* of and *feelings* towards the language, the GCSE course, and their own abilities. This complexity suggests that a broader understanding of learners’ CS concepts, which includes affective features and attitudes towards the learning process, may also need to be incorporated into future operationalisations of this construct. Each of these themes is discussed in turn.

4.2.6.1. Self-Assessment of Ability

In their discussions of their current confidence levels, learners would refer to assessments of their ability and performance, often using this as grounds for their degree of confidence. Of the 186 references coded under Current Confidence, 148 took the form of comments on participants’ overall performance in the subject, ‘I’m doing well’, or conversely, ‘I don’t understand Arabic at all.’ 26% of these references also

mentioned a specific skill or language component on which participants based their self-assessment:²²

Yeah, for me, reading is my weak point and then listening is my strong point. And then that balances out my grade. So... I feel okay when I get the grade, like, balanced out- Because... If I was to just to get my reading grade, I will not be happy with it... But because my listening grade pulls it up, it kind of balances the whole situation.

This breakdown of language skill assessments reflects both the composition of the GCSE curriculum, which is divided into separate *Reading, Writing, Speaking, and Listening* papers. It also hints at the overall assessment of learners' language ability being potentially misleading, since it may, as the interviewee above notes, represent an average of skill levels which may individually diverge quite significantly from each other.

What is also clear, and perhaps unavoidable, is the underlying reliance on grades as the primary point of reference on which to base measures of self-assessment, *'I feel like I could get, like, an A, if I try really hard.'* The question is raised as to whether basing responses on an anticipated or predicted grade actually constitutes a legitimate self-assessment. This point is made particularly acute by some participants' reflections on how their own grade expectations, based on their study efforts, are not necessarily matched by the grades they eventually receive.

Arabic is that subject, like... you just... you have to put in the effort and just have faith in God. That's how I see Arabic as. For the other subjects, you know what you have to write, you have to write that. If you don't write that, you've not got the mark. Whereas Arabic, it's either you do that, you have faith in God and just wish for the best. That's how I feel like.

This, of course, can work conversely:

Sometimes, when I think, like, I've flopped a test, but then my result comes back and I realise, like, I can actually do some of this! And it increases my confidence-Cos I'm, like, oh right, I actually know how to do this, I just need to practise it more. So, then I realise I'm not as bad as I think I am, I just need to do a lot more of my own work at home...

²² A number of data-points contained references to both general levels of confidence and skill-specific levels.

The ability component of learners' current-selves is therefore problematic given the ingrained GCSE-specific framework of grades upon which learners base their assessments, both of their current and predicted levels of ability, and the generalisation of skill level which such a grade forces them to make.

4.2.6.1.1. *Confidence*

Amongst students who indicated that they were confident with Arabic currently, a number of modifiers were repeated, including 'kind of', 'semi', 'quite' confident. This suggests a very cautious assessment of learners' own views of their grasp of the language. Only two participants of the 95 responded to this question by saying that they were currently 'very' confident with Arabic.

In contrast, 17 responses to the same question included varieties of the phrase 'not confident'.

I'm not comfortable- I'm not, like, confident at all... So, like, I'll probably drop it... Nothing makes sense to me, like, it's the words... Nothing. I just don't understand it. Like, at all...

Responses such as these indicate that a lack of confidence is precipitated by low levels of performance, often reflected in learners' grades. When elaborating on the reasons for their lack of confidence, many learners referred to their sense of feeling out of their depth with the language, particularly since starting the GCSE:

On a scale of 1 to 10, like, a 3 or 2.... Cos I don't know the basics...

... Since, like, Year 10... I've been flopping at Arabic. Like, I actually don't get it, at all...

Linked to comments on learners' self-assessments, comments on confidence were also often presented as skill-contingent, corroborating the necessity for domain-specific measures of confidence, rather than overall measures.

I'm not confident in speaking mostly... I think, um, reading it [...] I'm more confident, but writing, not as much.

There was no clear consensus on which skills were the hardest or inspired the most or least confidence. Of those students who did mention skill-specific confidence ratings, 9 mentioned speaking as an area of low confidence, whereas 6 mentioned reading as their stronger skill.

These findings corroborate those from the quantitative survey data analysis in the sense that they highlight the danger in conflating skills when measuring motivation. Whilst learners do not refer to MSA specifically in their discussions on their current selves, their generally low levels of confidence when discussing their GCSE performance could provide an explanation as to why MSA emerged as consistently significantly lower than other varieties in the quantitative findings.

4.2.6.1.2. *Perceptions*

A less direct reflection of learners' CS concepts can be discerned from their statements regarding their perceptions of the language and their current course of study. When asked about their experiences with Arabic, two dominant themes emerged. The first concerns the *difficulty* of Arabic (both as a language and as a GCSE subject), and the second concerns their perceptions of the value, utility and accessibility of the *GCSE specification*. As a reminder, this data was collected in the first year of the implementation of the new 9-1 GCSE examination system for Arabic and other subjects.

Following the first exploratory reading of the data, it was evident that *difficulty* was a recurrent theme. References to Arabic being 'difficult' or 'hard' were coded under this *in vivo* heading. Across all interviews, 119 references were coded under the category *Difficult*, with the word 'difficult' itself used 85 times in reference either to Arabic, or specific skills, across all interviews. The challenge with this particular section of the data was differentiating between Arabic the language and Arabic the subject. Some references make this distinction clear: '*Definitely, probably, one of the hard subjects that I'm learning*'; '*I feel like Arabic would be one the hardest GCSEs that I'm going to do out of all of them*'; others, however, appear to conflate the two: '*It's interesting, but I found it difficult. And-- Yeah, it was hard, and I don't feel very confident in Arabic*'. Therefore,

recourse to the context of the reference is necessary to determine, where possible, what was being referred to.

4.2.6.1.2.1. Overall Difficulty

What is interesting to note is that multiple references to the difficulty of Arabic invoke comparisons to previous years of study.

In Year 7 and 8, it was easier back then. But when we started the GCSE, then I found it more harder [sic]. So, I'm struggling a bit [more] with it than I was before.

This raises the vital issue of contextualisation of learners' CS concepts. Though Arabic can be said to be a challenging language for non-heritage learners, particularly for those who speak more distantly related languages as their L1s (Stevens, 2006), the difficulty many of the participants refer to here is a temporally comparative one. In describing their experience with Arabic, 29 students referred to the contrast between learning Arabic in KS3, which they described as 'easier', and the GCSE. It was noted that there was a 'big jump', between these two learning contexts in terms of content and pace of learning.

... It was a massive jump from Year 7, Year 8, Year 9 [KS3] to all the way to Year 10. So learning really slowly at a good pace and we were enjoying it. And then all of a sudden, Year 10 comes and everything is, like we had to learn in, like, four months. And because of that, it was really stressful.'

This perception of difficulty, whilst not an explicit reference to ability or confidence, indirectly indicates a level of discomfort with learning the language, enhanced by the contrast between previous years of study and the GCSE. This contextual grounding of descriptions of CS concepts was the first indication of the necessity of incorporating measures of learners' Past Selves into the motivation framework. This is discussed in detail in Section 5.2.4.

4.2.6.1.2.2. Language Difficulty

References to the difficulty of Arabic as a language fell into two themes: grammar and comparison to other languages. Grammar, 'the hardest bit of Arabic', was highlighted as

a major factor contributing to the difficulty of the language. Where this was expanded upon, verb tenses were mentioned as an area of difficulty, with one student who had studied these systematically in the mosque admitting to the desire to teach these to the class herself, according to the method she had used to acquire them, having seen how much her classmates struggle in this particular area. This admission seems to suggest that part of the problem with learners' lack of comprehension of tenses is the teaching approach, and/or the lack of basic knowledge required before being able to internalise this new language information.

A further suggestion of this disrupted learning trajectory is the fact that some participants expressed a desire to be learning a European language, such as French or Spanish, or Urdu, since they considered these to be '*easier*'. One participant, for example, expressed that French was easier given that they had studied the language since primary school and that it shares a script with English: '*it was more of, like, English-based subject, it was more easy to understand*'. A preference for Urdu is explained by the fact that it is a common first language amongst the sample.

4.2.6.1.2.3. GCSE Issues

As part of the Creative Multilingualism project into the new GCSE specification, a recent report by the researcher revealed the unreasonable demands in particular for vocabulary knowledge placed upon students by the new examinations (Ramezanzadeh, in preparation). References were made by students to the volume of vocabulary needed as evidence of the difficulty of the 9-1 GCSE specification, one of whom dubbed Arabic '*the language that has, like, a 100-page booklet.*'

Student 1: ... it's the new words and... they gave a specification out saying that these 300+ words, gonna have to memorise in Year 11.

Student 2: It's about a hundred pages we have to memorise.

Student 3: A hundred pages of words in Arabic. So that, when Mr [teacher] told us that, that was when my confidence kind of dropped.

Student ?: Since this 9-1, and then he [teacher] wants us to, like, learn 300 words over the summer holidays... So obviously there's a pressure in learning

Arabic, because the old GCSEs, it was... very simple. People in our class in Year 8 and Year 9 could have taken that and still got an A, an A, or a B. But since it's too late now...*

Researcher: Too late?

Student 3: Too late to do the old one that they did this year. Uh... We have to do the 9-1 and the 9-1 is very, very, very rigorous subject.

The above excerpt also mentions a key factor in the construction of learners' perceptions towards the current GCSE specification, the issue of comparison. Reinforcing their views of the difficulty of the content in terms of workload is the experience and/or understanding that the previous specification was less demanding. Some students mentioned that their older siblings had sat the exams prior to the changes and had found it easier. One cohort in the sample (n=7) had actually been entered early for the GCSE in Year 10 to benefit from the apparent ease of the earlier examination system.

References to the difficulty of the 9-1 specification are replete within the data: 125 references in total were coded, from a total of 35 of the participants. Several students spoke of the unfairness of the new requirements, though most interestingly in the context of this study, many reported the impact that it had had on their confidence and expectations of success. This feeling was sometimes addressed by working on exam preparation techniques, but in general the perception of the new exams as being '*much harder than the old ones*' had a negative impact on learners' confidence levels and their anticipated grades. As shall be addressed in subsequent sections, this can also have a significant knock-on effect on students' IS concepts: '*Since I've heard about, like, how they've changed the stuff in the new Arabic, I don't see myself achieving that goal- and becoming what I want to be.*'

Reflecting on the pervasive theme of comparison as a constructive device for learners' perceptions of Arabic and the common theme of difficulty formulated on this basis, what is interesting to note is the sensitivity needed when examining a participants' CS concept to the object *vis-a-vis* which they are describing themselves. In this case, descriptions of Arabic as 'hard', and the subsequent impact of this view upon learners' confidence do reflect an experience of the language, and, to a greater degree, outline

their experience with a course of study. The language did not get ‘harder’; rather, the course requirements became seemingly more demanding. The conflation of language and subject obscures both the educational context—both current and previous—as well as learners’ other individual self-concept references. Such nuances should be reflected in any instruments designed to measure the intricacies of learners’ self-concepts.

Returning to the quantitative data here provides evidence for this point. In the survey, learners were asked about their feelings towards the GCSE and towards Arabic as a language. A paired samples t-test indicates that there were significant differences between learners’ feelings towards Arabic and the GCSE, with the scores for Arabic being significantly higher at both timepoints ($M_{T1}=0.395$, 95% CI [0.271, 0.52], $t(262)=6.261$, $p<.001$, $d=.381$; $M_{T2}=0.497$, 95% CI [0.335, 0.659], $t(172)=6.045$, $p<.001$, $d=.462$).

4.2.6.2. Affective States

Current self-concepts do not exist within a vacuum. Not only are they constructed using situational and temporal references, but they also come attached to affective states. The presentation of a person’s self-rated ability is invariably intertwined with emotions concerning it. The nature of learners’ self-concepts therefore, at the very least, needs to acknowledge this affective dimension. In their specific context, an assessment of one’s ability as ‘quite good’ might reflect pleasant surprise, improvement, or disappointment. Perhaps more so than learners’ explicit statements concerning their confidence or abilities, these emotional indicators paint a more detailed picture of the nature of learners’ Current Selves. A number of these emotions were pervasive across the sample, some of which have been hitherto alluded to. Positive affective responses revolved largely around enjoyment of Arabic and a sense of accomplishment. The most prominent negative affective state was coded as despondency. Overall, references to negative feelings ($n=122$) were more frequent than positive ones ($n=81$).

4.2.6.2.1. *Negative Affective States*

4.2.6.2.1.1. *Despondency*

Despondency was isolated from amongst the primary coding of negative emotional states, which also included feelings of frustration and pressure, to refer specifically to a state of hopelessness, low spirits, or dejection as a result of the experience of learning Arabic. There were 56 references to this feeling throughout the interviews. Students who presented despondent responses often did so in reference to their expected grades or to their current perceptions of their abilities. While only used a total of eight times by four different participants, the word ‘flop’ was a particularly emotive term used to describe this feeling.

But inside I'm not that confident, to be honest, but outside I may seem confident in the class and stuff but inside I know that I'm gonna flop.

Though perhaps not consciously on the part of the interviewee, this word invokes the image of a dive. Whether or not this was intentional, a dive—a feat which takes effort, precision, focus and courage which ultimately ends in a ‘flop’, an embarrassing, possibly painful result—serves as a particularly powerful metaphor for the experience which some learners describe in relation to their Arabic studies.

As a broader category, however, the *despondency* code incorporated feelings of a sense of hopelessness, not necessarily associated with grades. That is, it referred to a state in which learners were either overwhelmed by the perceived difficulty of the content or feel so alienated by it that they give up.

I feel like I don't understand Arabic at all, in lesson I'm just sat there and when [Teacher] asks me a question, I'm just sat there... everyone's just staring at me and I don't understand it at all...

Student: ...from Year 7 to now, I got better. But I'm not that good... I just sleep in Arabic.

Researcher: What puts you to sleep?

Student: [Teacher] goes too fast... And I like, I just give up.

This particular sense of hopelessness is exacerbated by the mismatch between effort expended and results achieved, as noted by some students.

... Sometimes I feel like I tried so hard... I don't get the grade I want. So then... it puts me off.

This frustration is brought into focus, again, through the comparison of studying Arabic to the experience of studying other subjects, in which it appears that effort is translated more transparently into results.

Honestly, I don't feel like I can pass my GCSE exams for Arabic. Like, I feel like I want to pass– But I'm not sure how... Cos... I don't know how to revise. Like, I try learning my vocab but then, like, none of the vocab that I learnt comes up on the exam.

Arabic is that subject, like... you just... you have to put in the effort and just have faith in God. That's how I see Arabic as. For the other subjects, you know what you have to write, you have to write that. If you don't write that, you've not got the mark. Whereas Arabic, it's either you do that, you have faith in God and just wish for the best.

This final quote, also used in reference to perceptions of the GCSE, is particularly noteworthy as it incorporates a religious dimension of 'surrendering' one's will to God having expended effort towards a goal. This participant should perhaps be commended for their attempt to repackage a vexing situation into a positive faith experience, since the alternative appears to be mere frustration. However, the underlying implication of the quote is the apparent futility of learners' efforts when it comes to doing well in Arabic and the lack of transparency in the rewarding of effort, which in the simplest usage of the term can only be described as incredibly demotivating.²³

A consequence of learners' negative experiences with Arabic was often the de-

²³ The role of God in motivation and achieving success is a fascinating avenue for further exploration. Whilst this was beyond the purview of this study, an investigation of concepts of the divine through the prism of attribution theory (Weiner, 1986) presents a particularly interesting area of focus. The student in the quotation, for example, could also be interpreted as attributing the remainder of the formula for success to God (an uncontrollable, presumably stable, external (or perhaps not) locus of control, having expended her own efforts in the equation. Her statement, thus, potentially represents less helplessness than a worldview in which future events are not solely contingent on her own actions, nor even those of other human beings. Perceptions of ontology, theology, and indeed teleology all potentially overlap and play a role in the pursuit of goals amongst participants from religious backgrounds.

prioritization of the subject in favour of other ‘core’ GCSEs, such as English, Maths, and Science. This appears to have been precipitated by: the lack of clarity as to the translation of effort into grades through study and revision; the perceived difficulty of the subject; and existing negative affective states often associated with Arabic. As a result of these factors, learners reported their neglect of the subject, which inevitably leads to a vicious cycle of frustration, despondency, and underachievement:

You really have to think about it, and then you have other subjects that you have to worry about... And then Arabic comes in the way.

4.2.6.2.2. Positive Affective States

4.2.6.2.2.1. Enjoyment

There were 25 mentions of enjoyment in relation to Arabic, some of which incorporated expressions of interest in specific areas of the subject:

I think that Arabic, even though it's one of the hardest languages, it's really interesting to learn, because like all the root words– where one word comes from another, it's just really, like, interesting. And obviously it plays a big role in my religion as well– So... that's all the factors, but, yeah, I think it's really interesting to learn as a subject.

A number of students expressed an interest in the linguistic structure of the Arabic language, particularly its expansive root system, as well as in the novelty of learning a new, less common, or ‘hard’ subject. In some cases, the structure and unique features of Arabic, in comparison to other languages students had learned, inspired a sense of enjoyment in studying the language: ‘*I really enjoy Arabic, because it's, like, more structured than other languages.*’ Other participants, particularly those in one of the classes, spoke of enjoyment and interest in the subject, inspired by the teacher and lessons:

... Once you come in Arabic everyone is quiet, because they actually want to learn it because [Teacher] makes it, Arabic, sound so interesting and enjoyable to actually learn...

4.2.6.2.2. Accomplishment

Feelings of accomplishment in the successful application of knowledge—whether inside or outside the classroom—were evidence of positive emotions associated with learners' Current Selves. Although these concrete experiences and feelings of accomplishment in the present were sparse within the data, they appear to provide an anchor for learners' confidence and thus a potential source of encouragement which can fuel the desire to continue with the path of study.

... If you read the Quran and you understand, like, I don't know, maybe half a sentence or a sentence, you're... so, like, accomplished. [...] Yeah, you just feel so smart and it's just so nice.

Overall, although both positive and negative affective statements were present, the disparity between positive statements (39.9%) regarding learners' current experiences with Arabic and negative ones (60.1%) is salient. The latter were expressed, in particular, in reference to the experience of studying the GCSE course. The disproportion in positive and negative emotions related to Arabic—both as a language and a GCSE subject—presents a worrying area for practitioners and corroborates earlier findings of the researcher (Ramezanzadeh, 2015). The various nuances captured by the qualitative analysis of the interview data also provide some explanation for the persistently low MSA CS scores (and scores for other varieties) in the survey, despite the longevity of learners' study of the language.

4.2.6.3. Gender

Whilst there were no significant differences between the genders' CS concepts at either timepoint in the survey data, a number of interesting observations can be made when exploring the affective domain of the CS in relation to gender amongst interview participants. Independent samples t-tests comparing males and females' feelings towards Arabic and the GCSE, measured by two Likert-scale items in the surveys, showed no significant differences at T1 and T2. However, the interview data suggests there may be some underlying differences when feelings are analysed in a more focused manner. Figure 4.10 presents the frequencies of references to different affective states amongst males and females. The bars represent the percentage of participants within

each gender who expressed references to the different affective states.²⁴

One such finding, in particular, stands out. The highest number of references made amongst male students in relation to their feelings towards learning Arabic were in reference to overall positive feelings, whereas amongst females this was despondency. That is not to say that female students did not experience positive feelings, this was in fact the second highest category. What this suggests is that the views of female students were more varied in terms of their positivity or negativity, whereas males were predominantly positive.

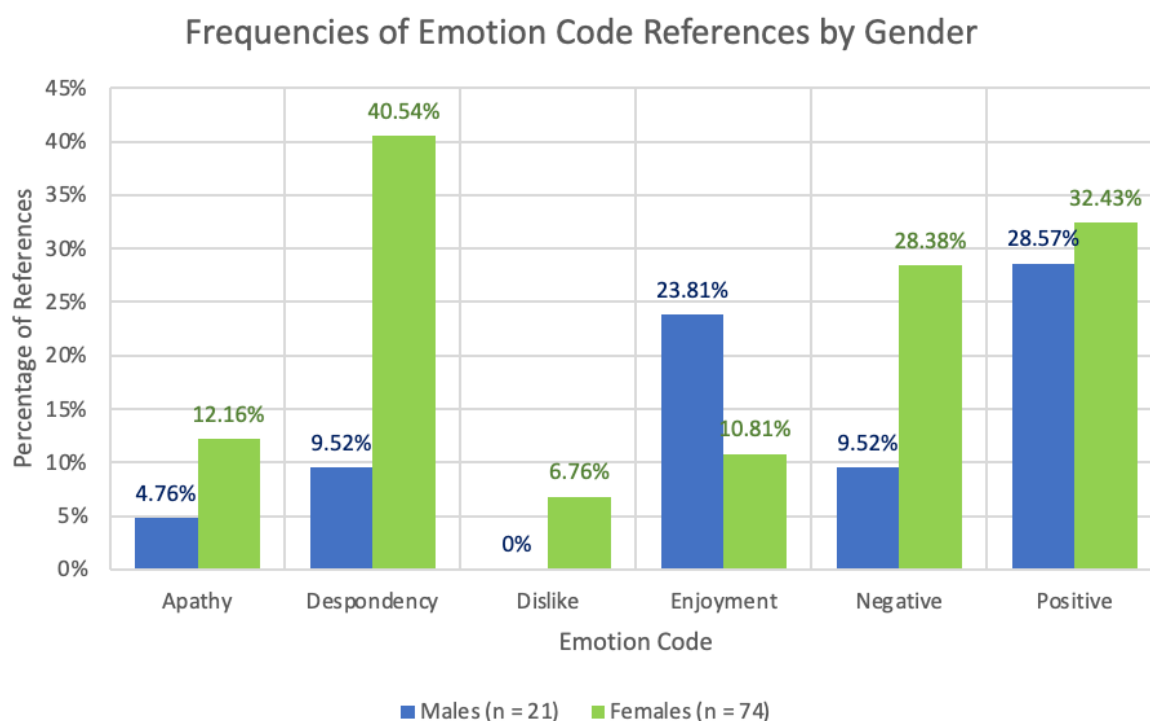


Figure 4.10. Bar graph for references to Emotions codes by gender

The contrast between these observations and the findings in the statistical data with regards to feelings could be explained by a school effect. The two classes from which male students were drawn both mentioned favourably their learning experiences. In the former, learners made reference to quizzes and activities they enjoyed, as well as a whole lesson the teacher had dedicated to discussing possible future uses of Arabic. In

²⁴ Since participants expressed multiple emotions in regard to their learning of Arabic, these percentages do not total 100% for each gender. This also applies to the graph for heritage and non-heritage learners.

the second class, learners described their enjoyment of the subject and experiences studying other languages, all of which, it could be argued, contribute to the overall positive emotions expressed towards Arabic in the interviews. This possible evidence of a school-level effect is discussed in Section 4.4.5.2.

It is necessary to note that sample sizes are small, and the male groups was less than a third of the size of the females. Since gender overlapped with classroom groups this also constitutes a limitation as to the interpretation of these results. However, whilst these observations cannot be said to be robust findings, they do highlight the potential for further research in this area.

4.2.6.4. Heritage Learners

In contrast to the absence of significant differences between genders in the survey data relating to the CS, HL scores emerged as significantly higher than NHLs at T2 (see Section 4.2.3).

In addition, a paired samples t-test indicates that at T2 HLs' feelings towards Arabic and the GCSE were both higher than their NHL counterparts ($M_{Arabic} = -.865$, 95% CI [-1.328, -0.403], $t(95) = -3.714$, $p < .001$, $d = .61$; $M_{GCSE} = -.888$, 95% CI [-1.413, -.363], $t(172) = -3.340$, $p = .001$, $d = .581$). The effect sizes for both were medium, indicating that the heritage background factor played a relatively strong role in this difference.

This appears to be corroborated by the interview data, which shows higher scores for HL learners in the *enjoyment* domain, in contrast to NHLs who demonstrate higher levels of despondency, as illustrated in Figure 4.11.

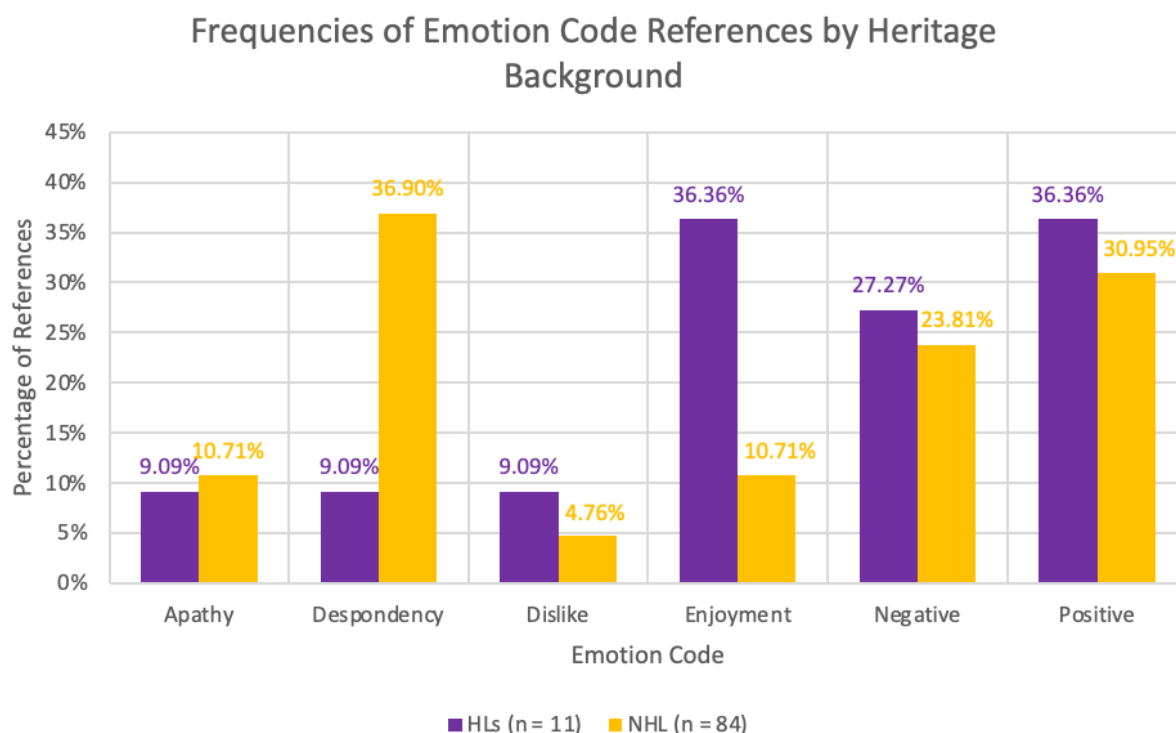


Figure 4.11. Bar graph for references to Emotions codes by heritage background

As previously noted, since the sample sizes are limited, further research is needed to validate these findings. However, the congruence between the findings from the three sets of data (i.e., Overall CS scores; feelings survey scores; and interview data) suggest that HL do exhibit more positive affective states in relation to learning Arabic in schools. Unlike the gender category, which is split along school lines, HLs were not concentrated in particular schools, thus indicating that it is likely to be their background which lends itself to these differences.

4.2.6.5. Complexity of the Current Self—‘But...’

Despite the obvious variation in emotional responses or interactions with Arabic, it would be misleading to present positive and negative affective expressions solely as a dichotomy, a bipolar scale along which students can be neatly categorised. Evidence in dispute of such a scale is present in the frequent use of the word ‘but’ in learners’ elaborations of their Current Selves. In total, there were 44 such statements which used the word ‘but’, or a synonym, to express contradictory feelings towards Arabic or perceptions of their motivational self-concepts.

It's interesting, but I found it difficult.

I'd just love to learn Arabic, but it's really hard.

I do want to learn it, but I know I'm not going to learn the Arabic language if I do my GCSE.

We don't hate Arabic... It's just hard...

I proper really wanted to do Arabic, but it's just gotten really hard.

I want to pass my GCSEs but it's just so boring. Like, really boring.

The plethora of juxtaposed statements alerted the researcher to the complexity of learners' CS concepts. It is entirely possible for learners to 'love the language' but at the same time be very frustrated with the learning process to such a degree that they are ready to give up. It is also possible that the formula *positive statement + but + negative statement* (the reverse formula was four times less likely to appear in the data) was provided in response to the presence of the researcher, and a fear of being seen as overly negative from the outset of the interview, particularly since the participants were aware that the researcher was focused on Arabic and had learned the language herself. It is also possible, given the religious associations of Arabic, that learners were hesitant to criticise or complain about the language in the presence of a Muslim adult. The role of emotions in learners' CS concepts thus constitutes an interesting area of focus, and is discussed in the next chapter (see Section 5.2.1.4.4).

4.2.7. Ideal Self

Participants were asked how they saw themselves using Arabic in the future. Since there is some debate as to the capacity of teenagers to have developed future self-concepts (Dörnyei, 2009), learners were first asked to imagine themselves 5-10 years in the future and to describe their vision. The purpose of this question was two-fold: a) to encourage learners to imagine themselves in the future before inviting them to describe what role Arabic plays in that image, if any, and b) to determine whether the lack of an Arabic future self meant an absence of a general future self (FS) image or the ability to generate one, or if it was indeed an Arabic-specific component that was missing from participants' self-

concepts. There appeared to be no link between the provision or absence of an Arabic-specific FS and a general FS image, indicating not only that learners of this age group are capable of generating IS images and have given thought to these, but also that they were also conscious of whether or not Arabic played a role in them.

Based on their responses to the Arabic future self (FS) image question, reported Arabic-specific IS-images were organised into the following seven categories: *Family; Travel; Skill; Job; GCSE-oriented; Religion; and No Image*. Frequencies of interviewees who made reference to each category are provided in Figure 4.12. As is apparent from the percentages, the most popular of these FS images was Travel, followed by Religion. The Skill category, in which learners expressed their desire to be able to communicate in the language, without reference to a specific context of use, was also frequent. This was followed by the No Arabic category, which involved over a quarter of participants, who expressed explicitly that they did not see themselves using the language in the future. Whilst each of these categories will be analysed in turn, it is worth noting that many participants included multiple categories in their answers.

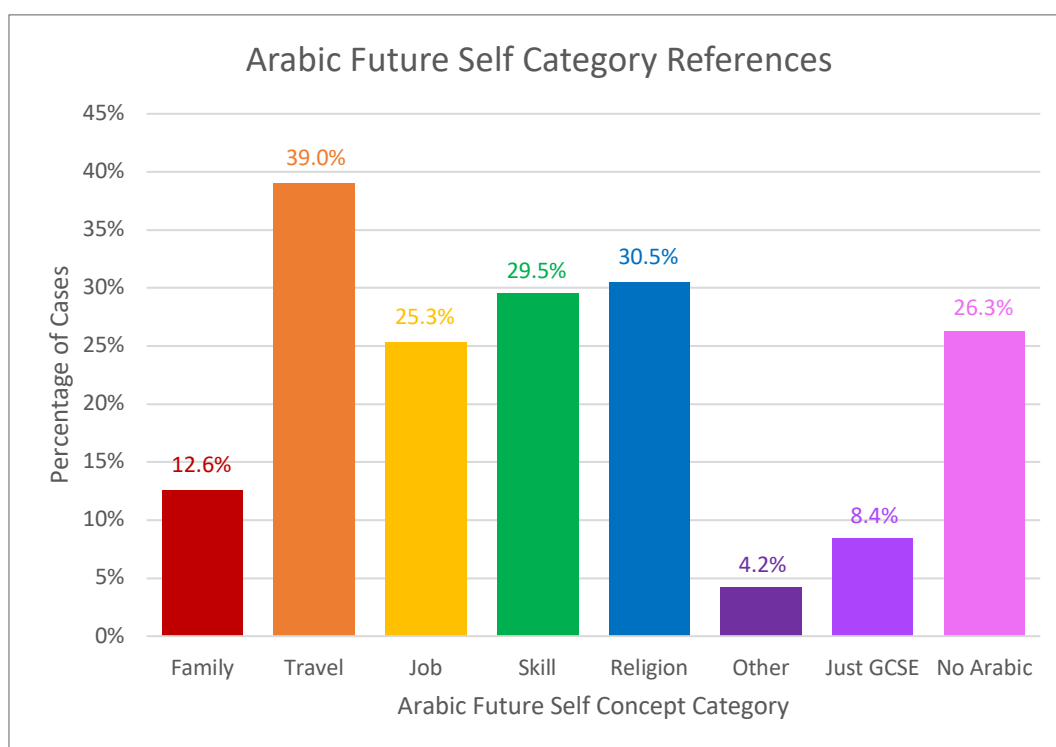


Figure 4.12. Bar graph for references to Arabic FS images

4.2.7.1. Family

Participants who invoked communicating with their Arabic-speaking family as a component of their IS concept did so in three different ways: to communicate with their immediate, Arabic-speaking family at home; to speak to family living in Arabic-speaking countries; and to visit relatives in Arabic-speaking countries.

Whereas the second two categories involved learners from diverse ethnic backgrounds whose family members either hailed from or had moved to Arabic-speaking countries, the first self-image was exclusively expressed by learners in the HL category, whose families spoke Arabic at home. In most cases, these learners' families spoke colloquial Arabic:

I think dialect Arabic I'll probably use more with my family, because none of them speak, like, fuṣḥā [standard Arabic].

This would also be the case with Arab family members living abroad, since the dialect is the primary means of communication and the L1 of native Arabic speakers, though there is the possibility for those who are NNSs living abroad to have learned MSA before becoming proficient in the dialect.

HLs who cited communicating with family as a component in their FS image make reference, as in the above quote, to the different types of Arabic used at home and in school, raising the issue, directly or indirectly, of the utility or at least transferability of 'school' Arabic to their intended goals.

4.2.7.2. Travel

Travel is commonly cited as a motivation for language learning (Clément, Dörnyei & Noels, 1994; Graham, Courtney, Tonkyn, et al., 2016; Taha, 2007) and this study was no exception. *Travel* emerged as the most frequently referenced Arabic FS image in the sample. It is important to note that travel in this study consisted of two types, namely: travel for tourism, and living abroad.

Multiple participants spoke of the possibility of travelling abroad to Arabic-speaking countries and using Arabic to communicate. This included travel for tourism purposes, to places such as Dubai or Egypt; and religious purposes, specifically to Saudi Arabia in order to perform pilgrimage.

Similar to the use of Arabic with family members, this again raises the issue of Arabic varieties and the conflation of dialects and Modern Standard Arabic under the broad banner of 'Arabic'. This is not to say MSA would be useless in an Arab country for the purposes of communication; indeed, some learners mentioned past experiences of having been able to partially communicate using the language they had acquired in the classroom abroad. Whilst the issue of language type may not be as pressing in the context of holidays, or even pilgrimage which is usually performed in groups (and which is well signposted), the sole use of MSA for the purpose of living abroad is more problematic.

I definitely want to speak Arabic. Like, I'll move to an Arab country for definite. I've always dreamt of that, like, in Year 7, I would say, "Oh, I want to speak fluent Arabic." So, like, it was always my dream to speak Arabic.

What constitutes 'fluency' in Arabic is a contentious topic. Answers usually revert to a highly contextualised definition. What is unavoidable in the context of living abroad is that knowledge of the local dialect is a vital component, which the GCSE does not directly facilitate, nor explicitly include.

Upon closer inspection of participants' travel-oriented responses, it became apparent that there was a distinction between answers that alluded to pre-established thoughts of travel abroad, and those which seemed to be provided as mere hypotheticals. Contrast, for example, the following two quotes:

I would see myself using Arabic because... out of most countries that I've travelled to and would travel to are mostly, they speak Arabic. And communicating with the people would need the use of the language.

Only if I meet someone, like if I went to an Arab country then I'll maybe use it then...

Both reference travel as a component of an Arabic-related IS. However, in the first, the participant begins their response with the words *'I would see myself using Arabic'*, a clear statement of a FS concept, and then goes on to reference past experiences upon which they are basing this, ending with a justification. The second introduces the travel component with *'only if'*, a restrictive, hypothetical statement, further qualified by *'maybe'*, suggesting that even if they did go abroad (which they may not), there is a chance they still may not use the language. A re-examination of the travel-oriented responses indicated that this use of hypothetical statements as responses to questions concerning FS images was common. Indicators of this type of response included *'only if'*; *'if I/ you'* (the *'you'* invoking further distance between the participant and the hypothetical scenario); *'maybe'*; and even statements where the FS is immediately undermined by a negation of the scenario: *'if I think about it I just use iTranslate, or Google Translate...'*; *'Maybe if I want to go to live in an Arab country, but I don't think I will.'* In total, there were 22 uses of these explicit hypotheticals, 13 of which were in reference to *Travel*.

Several participants also used follow-up questions about the vividness and frequency of their Arabic FS concepts as an opportunity to disclose, indirectly, that these references were hypothetical and that they did not give them much thought. According to the L2MSS, in order for an IS image to be motivating, it needs to be desired, vivid and frequently activated (Dörnyei, 2009). Clearly, the hypothetical statements do not meet these criteria. This data highlights the possibility of learners' being able to conjure up pre-fabricated, generic answers to questions regarding future Arabic usage when requested, which, in cases where no further details or information about these FS concepts are elicited, may skew the data quite profoundly. Whilst the provision of hypothetical scenarios rather than established FS images was not done by all participants, this subtle difference highlights the necessity of developing instruments sensitive enough to be able to distinguish between the two types of responses.

4.2.7.3. Job

The use of language for a specific career or job was also a commonly cited component of learners' FS images, as has been reported in other Arabic studies (Brosh, 2013; Winke &

Weger-Guntharp, 2003). In this sample, the code *Job* covered four categories: *Arabic-centred job*; *non-Arabic centred job*; *work in an Arab country*; and *generic Arabic jobs*.

Very few students mentioned Arabic-focused jobs as constituting a component of their IS concept. Jobs that were mentioned included: translator; interpreter; and teacher of Arabic or Islamic Studies.

Non-Arabic-centred jobs covered those in which the use of Arabic would be seen as a useful, but not primary skill. Learners remarked on the possible benefits to doctors, lawyers, and social workers of speaking Arabic:

I wouldn't use it in my job exactly, but maybe if I had Arab patients or something, then I'd make them feel more comfortable if I speak Arabic with them.

Going to work in the Middle East did not usually involve descriptions of a specific job, although one student did speak of wanting to open oil reserves in Saudi Arabia, but nevertheless was included under the heading of *Jobs* since work was specifically mentioned in the description of plans or desires to move abroad.

In a similar vein to the hypothetical travel scenarios which some students provide, participants also appear to provide generic jobs in which Arabic can be used in response to the Arabic FS image question:

... It only comes in limited jobs. Like, if you want to be a translator or things like that, but I'm not interested in things like that.

Again, this quote represents a combination of indicators of a hypothetical or generic response, using 'only' or 'if you'; and the immediate undermining of the former part of the response, suggesting that participants are aware of such potential uses but have not integrated these into their own IS concept.

4.2.7.4. Skill

Rather than envisioning themselves in a specific job or travelling abroad, some students

chose to describe the particular skills they wished to gain in the future. When asked how they saw themselves using Arabic in the future, the majority of those in the Skill category described themselves speaking Arabic (coded *Skill: Communication*). This category involved communicating abroad and in the UK but was not tied specifically to travel or jobs. While some of these participants stipulated the desire to speak ‘fluent’ Arabic, others presented a FS image which involved a more limited delineation of their future skills: *‘Enough to, like, be able to communicate, like, can tell people what I need to say.’* Both would fall under the traditional categories of ‘communicating in Arabic’ in the future, but this difference in desired ability could be seen to represent a significant difference in the potential distance between participants’ Current and Ideal Selves.

In addition to describing the desire to communicate in Arabic, some participants went further and explained why they saw this skill as ‘an asset’. They made reference to the benefits of being seen to speak multiple languages, such as being seen as ‘knowledgeable’, and making them ‘stand out’ and ‘looks good’, particularly on a CV. Seven students made reference to their desire to study Arabic in the future, in order to enhance their skills, though none of them expressed interest in taking the subject at A Level or at university. Students who referred to future learning spoke of online courses, or ‘Alimah style programs.

4.2.7.5. Religion

The ability to understand the Quran was one of the most prominent religious motivations for studying Arabic. 62 references were made to religion as part of participants’ FS images, 42 of which involved explicit mention of the Quran. Of the 167 references to the role of religion in learning Arabic overall, 98 of these revolved around the Quran, demonstrating its significance in learners’ FS concepts and perceptions of the language from a religious perspective.

... Obviously cos I’m Muslim, I think I’ll be able to understand the Quran... even a little bit, just to kind of... help.

This prominence is predictable given the Muslim background of almost all participants, plus the predominance of Islamic schools in the sample and the importance given to

interacting with scripture in the Islamic faith. This categorisation of FS concepts under the banner *Religion: Quran* however, masks a number of nuances. The first is the distinction between *reading* and *comprehension*. Reading the Quran in Arabic is considered an act of devotion in the Muslim faith. Muslims often aim to complete an entire reading of the Quran during the sacred month of Ramadan, for example. However, what is actually being referred to by the term ‘reading’ the Quran is, in fact, phonological decoding (reading aloud). The integration of Arabic into the everyday lives of Muslims is alluded to at the beginning of the abovementioned quote through the use of the word ‘obviously’. The Quranic script includes diacritical markings unlike Modern Standard Arabic, making it easier to decode for native and non-native speakers.²⁵ Simply pronouncing the words of the Quran, without comprehension, is considered to be worthy of reward. Comprehension of the Quran is considered a separate, more advanced skill. Though it is often aspired to, it is not seen as intrinsic to one’s faith necessarily, though it is deemed to be of great benefit. The necessity of the comprehension of the Quran is further undermined by the proliferation of translations into multiple languages, often printed in parallel to the Arabic. Many students referred to the existence of translations as justification for their focus on, or satisfaction with, the goal of simply phonologically decoding the Quranic script.

When questioned as to the reason for the integration of understanding Quran into future self-concepts, those who had identified this particular skill answered with responses related to faith-enhancement desires.

Because if I understood what Allah [God] was saying in the Quran, then it’d probably have a good impact on my life, because I’d understand the message from Him, directly—because He revealed it—like, try to become more better in my faith. Like, it would strengthen it, so every time I read ... [or] I pray, if I understand what I’m saying, then I wouldn’t ... just read it and like, think about other things, because sometimes that happens. So, if I do focus more on the words and I understand what God is saying, then it makes me feel more connected [to God].

²⁵ By way of reminder, Modern Standard Arabic represents only the consonants and long vowels but omits the short vowels (represented through diacritical markings). The Quranic script, however, includes systematic diacritics to mark the short vowels (added in the 7-8th centuries) making it much more readily phonologically decodable without needing to understand the meaning (which of course is necessary to infer words from context). See Section 2.2.1 of the Literature Review.

Here, the comprehension of the Quran, as opposed to the mere ability to read aloud, is identified as facilitating both: an unmediated understanding of God's commandments which in turn promotes adherence to them, as well as concentration in prayer (*'I wouldn't (...) think about other things'*), known in Islamic terminology as *khushu'*, which assists in enhancing a spiritual state of connectedness to God.

Some students (n=3) took these ideas further and linked their desire to engage with Arabic and understand the Quran with the teleological goal of getting to heaven.

Student: Obviously to help me understand it [Quran] in different ways, so I could increase my knowledge.

Researcher: Why do you want to do that?

Student: Oh, so that I could get to paradise.

Researcher: How important is Arabic in getting to paradise?

Student: Oh, it's very important. As it's you know, the language of the Quran, and the language that Muḥammad spoke was Arabic... So, then that's why it's important. It's the language of Islam....

Followed to their natural conclusion, it could be assumed that most students who express a desire to learn Arabic in order to understand the Quran, independent of explicit externally imposed pressures, would be pursuing this goal for this teleological reason, given the worldview of Islam as integrating an afterlife. The desire to understand God's message, as the previous quote includes, and the desire to concentrate and be fully present in prayer, as well as having perceived everyday benefits, are usually presented in this ultimate-goal framework, which learners would unquestionably have been taught in Islamic schools if not in their own homes.

A further noteworthy point regarding learners' religion-related self-concepts is that often, when asked whether they saw themselves using Arabic in the future, learners did not report wishing to use the language religiously. However, when asked whether religion played any role in their Arabic learning, participants almost always remarked that they wanted to be able to read and/or understand religious texts or perform rituals in Arabic. Even when learners remarked that they did not see themselves using Arabic in

the future at all in response to the IS question, some then responded with this faith-focused desire to the subsequent religion-oriented question. This is illustrated in the excerpt below, the first quotation of which addresses the Arabic FS question, and the second as a response to the question on the role of religion in learning Arabic given by the same participant:

I don't think I'll be using Arabic in the future, and to be honest I don't really know a lot of Arab[s], and even if I come across an Arab, I don't think I'd really understand what they're saying.

I do want to know what the Quran means whilst I'm reading it.

What does this tell us about the relationship between religion and learners' FS concepts? There are a number of possible ways to interpret these findings. The first is to assume that learners, in response to the religion question, provide a generic, face-saving answer, in a sense paying lip-service to the expectation of being able and willing to engage with religious texts and practices in the future. This may certainly be the case for some students, and the face-saving approach may be made all the more acute by the overt religious affiliations of the researcher.

However, not all students provided this response; in fact, a number of participants from Muslim backgrounds (n=8) stated that they did not see religion as playing a role in their Arabic learning and did not provide a religious FS image. An alternative, though not contradictory explanation involves the fact that learners, to varying degrees, have already integrated this skill into their daily lives. Quranic Arabic, at least the ability to phonologically decode it, is already a well-established skill amongst many of these students (the mean scores for learners' self-rated proficiency in word reading with diacritics was 5.64 at T1 and 5.54 at T2, measured on a 6-point Likert-scale) and is thus conceptualised not as a future goal, but a component of their CS, which may or may not require development. To corroborate this, where religion was mentioned as part of an IS, this generally involved comprehension, rather than simply phonological decoding (reading aloud) skills, which learners felt were further removed from their current skill set. This is consistent with the view that in order for something to constitute a component of a FS image, there needs to be a perceptible gap between current and desired abilities, which is in line with the self-discrepancy component of the original

L2MSS theory.

Another facet of the discussion on the distinction between IS and CS concepts in this domain is made evident in the choice of language participants employed when speaking about their religious FSs. Participants oftentimes used the term ‘when’ to talk about the application of their Arabic skills to religious contexts:

Obviously, I think it's mainly when you read the Quran, when you understand it as well. That's sort of the main reason if you can learn Arabic.

It's nice, like, being able to translate stuff that... sometimes when you're reading the Quran, you can know words and it's nice to know that.

The use of ‘when’ implies a certainty of usage. In contrast to ‘if’ (mentioned above in relation to *Travel*), ‘when’ spans the tenses, and could be seen to represent an overlap of the past/present and future. ‘*When I read Quran*’ can refer both to something that is already being done, as well as to something that will happen in the future. This lack of a clear distinction between the present and future perhaps explains why this usage of Arabic was mentioned in response to the religion question, even when it was not mentioned as a component of learners’ IS concepts.

A further point of reflection, raised by the use of the phrase ‘obviously cos we’re Muslim’, and reiterated by a number of students is their perception of Arabic as a ‘foreign’ language. The perception of ‘different Arabics’ (see Section 4.2.5) and the prevalent theme of the difficulty of and frustrations with the GCSE, in contrast to the integration of religious-oriented Arabic, suggests that learners may not view Quranic Arabic as ‘foreign’.

Like, it's a familiar language to Muslims and it's written in the Quran... So it's quite, like, a special language.

I think with my family, everybody naturally learnt Arabic as they grew up, so for us, my whole family, they know how to read Arabic, which is, I think... for every Muslim so that they can learn Arabic, read Arabic, read with the Quran. But I think there was no pressure for me to, like, understand it.

Arabic is unique in the extent of its integration into the lives of non-native speakers of

the language, given its religious ties. The extent to which Arabic is considered 'foreign' is potentially an interesting area of future research, since the integration of a language into one's identity will surely have implications on both the desire and pressure to learn it (Jaspal & Coyle, 2010). Exploring these issues was, unfortunately, beyond the remit of this particular study.

4.2.7.6. Other

A small number of responses (n=3) did not fit into the seven categories listed above. These involved the desire to continue studying Arabic in the future, either as a main focus or as a subsidiary exploit alongside the study of religion. Some participants expressed the desire to completely start learning Arabic from scratch, *'I want to study it properly'*, either by learning online, taking courses, or by studying abroad.

Another student expressed the desire to teach her children Arabic in the future, whereas another expressed an ardent desire to marry an Arab. However, whether these fit into the category of *Family* as it was defined above is debatable, hence the choice was made to keep them separate.

4.2.7.7. No Arabic Ideal Self

25 of the 95 interviewees informed the researcher that they did not see themselves using Arabic in the future at all, although some later indicated that they did see themselves using Arabic to access Quranic texts. When asked to elaborate on their absence of an Arabic IS image three themes emerged: learners in whose FS images Arabic *Doesn't Feature*; those for whom Arabic had specifically *No Personal Utility*; and those who had previously imagined themselves using Arabic but for whom Arabic usage now seemed *Implausible*.

Statements such as: *'I think after, like, school's finished I don't think I'll ever use it again'* fell under the first category; i.e., learners who negated the possibility of using Arabic in the future but did not give reasons as to why. These statements ranged from expressions of doubt in future usage (e.g., *'There might be a slight chance but I doubt it'*) to complete

emphatic denials: *'No, not at all, no.'*

Those in the second category expressed the fact that they would not find Arabic useful for their futures: *'I feel like Arabic will get you nowhere.'* Others gave further reasoning, outlining potential jobs for which Arabic could be used but in which they were not personally interested:

... It only comes in limited jobs. Like, if you want to be a translator or things like that, but I'm not interested in things like that.

Replies in the third category represented the lack of likelihood of usage, either given the apparent inevitability of forgetting the language, or because learners had modified their FS images in response to difficulties in learning it:

I want to use it in the future because I want to travel when I'm older. But I feel like at the stage that I am right now, I feel like it's not gonna happen.

4.2.7.8. Gender

As with the CS, there were no statistically significant differences between males and females in the realms of the IS, based on the survey data.

When analysing the frequency of references to difference FS images in the survey data some noteworthy observations can be made. The highest categories in both genders included *Travel*, *Religion*, *Job*, and *Skill*, indicating a general consensus between the two groups as to the awareness of potential future uses of the language (see Figure 4.13). Interestingly, the category of Religion emerged as most popular for boys, whereas for girls this was Travel. The Family and Job categories were also more prevalent for boys. The latter can possibly be explained by school effect, since one of the boys' classes reported receiving a whole lesson on possible future uses of Arabic, whilst the former is explained by the fact that approximately a quarter of the boys' sample were HLs whereas only 8% of the girls' sample were from this group.

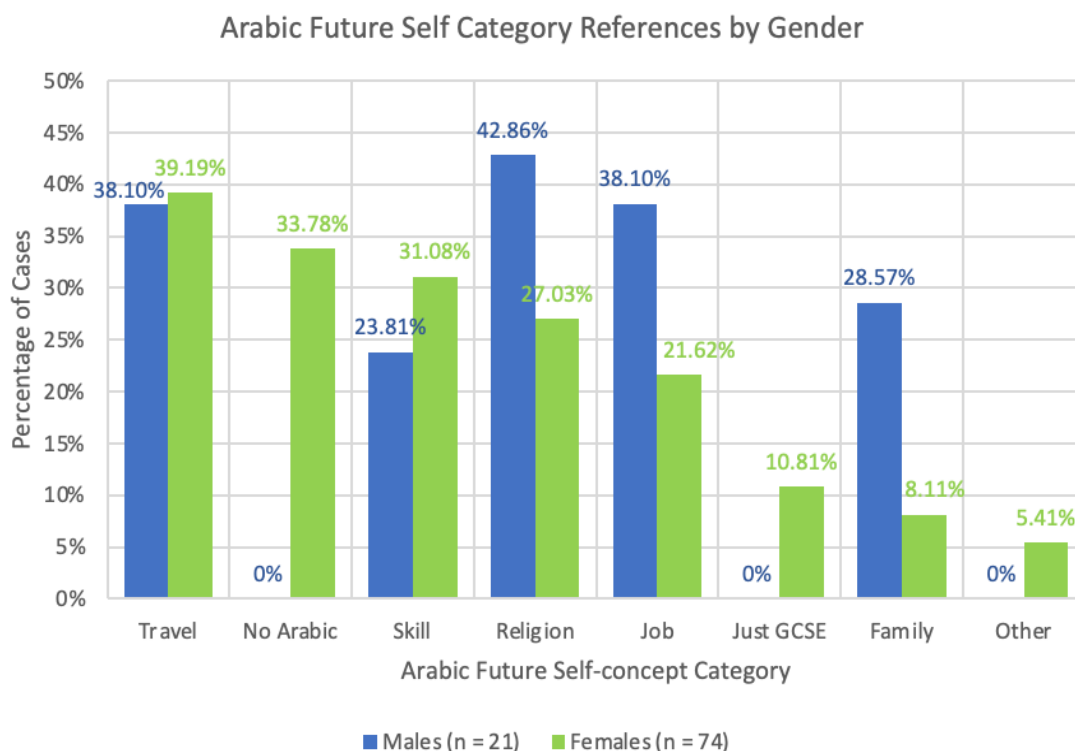


Figure 4.13. Bar graph for references to Arabic FS images by gender

More noticeable, however, is the high frequency of female students, over a third, who mentioned having *No Arabic* FS image at all. This stands in stark contrast to the fact that no male students expressed this view. As mentioned in the section on *No Arabic* above, this does not necessarily mean that learners did not later express a subsequent religious motivation. Regardless of whether this was the case, however, these figures are striking in an educational context. To return to the reference to school-level differences made in the section on Gender and CS, it is possible that the lack of the *No Arabic* category is explained by the nature of the two boys' classes participating in the interviews, one of which, as aforementioned, underwent a lesson dedicated to uses of Arabic in the future, and the other in which all students chose voluntarily to study the language, from which it could be inferred they had some form of FS concept in mind before even taking the course.

4.2.7.9. Heritage Learners

Heritage learners' raw IS scores emerged as consistently higher than NHLs' at both timepoints in the quantitative data. This is expected given HLs' greater exposure and

connections to the language. An exploration of the number of references to specific FS concepts in the interview data revealed a more detailed, though tentative, picture of this difference, as illustrated in Figure 4.14. Over half of the HL learners referenced their *Family* as the domain in which they saw themselves using Arabic, in comparison to only 7% of NHLs. Furthermore, they were over twice as likely to mention *Job*, and slightly more likely to mention *Travel*.

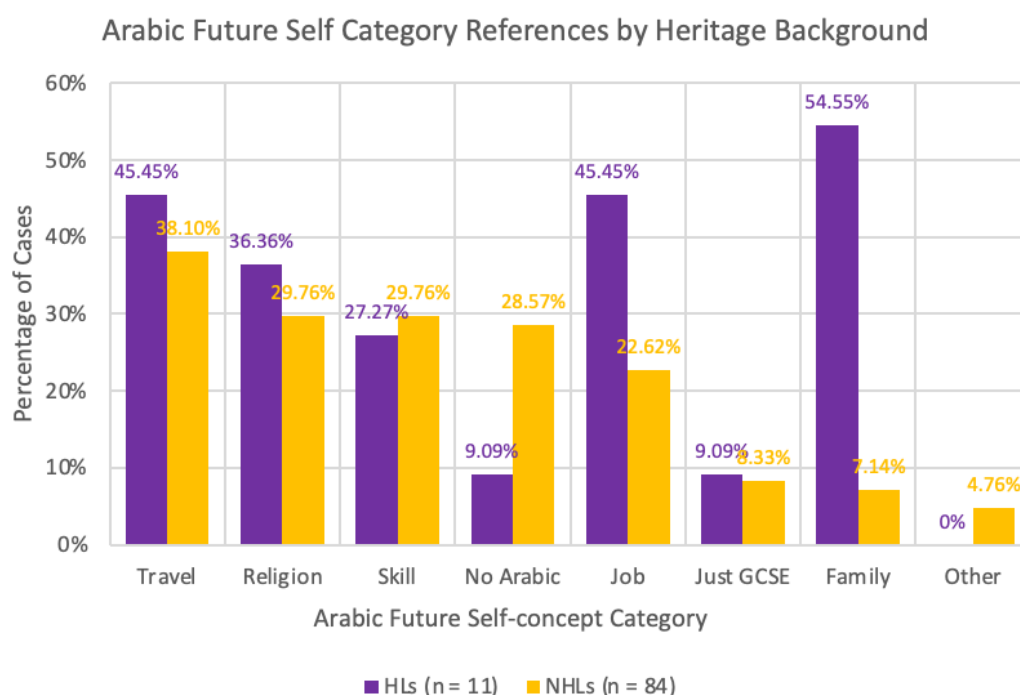


Figure 4.14. Bar graph for references to Arabic FS images by heritage background

In the NHL category, the most popular reference was to *Travel*, followed by *Skill* and *Religion*. One area of potential concern is the fact that 24 of the 84 NHL participants said that they did not see themselves using Arabic at all in the future. This fits with the CS picture painted in the interviews, in which 37% of NHL learners expressed feelings of *Despondency*. It could be hypothesised that the high levels of despondency (which could be interpreted as weaker CS concepts) could lend themselves to, if not correlate with, weaker IS concepts amongst learners, if indeed they have an IS concept at all. The relationship between motivational self-concepts will be examined in the path model analysis in Section 4.4.4.

4.2.7.10. 'If'

Returning to the sample as a whole, a recurrent theme across participants' answers concerning their FS images was the usage of hypothetical statements. Although mentioned above in reference to the *Travel* and *Job* categories, it is worth pausing at this point to discuss the category *in vivo* coded as '*If*', used by almost a quarter of students. Whilst there were certainly exceptions to the following, examination of these 'if' statements revealed three main categories of usage in the context of the IS: provision of hypothetical or generic IS concepts; contingent IS concepts; and less integrated or uncertain IS concepts.

The first category, discussed in the section on *Travel* (Section 4.2.7.2), involves the provision of generic answers to questions regarding participants' FS images.

It depends, like, if you're going to travel or if you're gonna, like, teach in a school.

Such 'if' statements are predominantly accompanied by references to travel or jobs such as Arabic teaching, provided as responses to Arabic FS questions. However, these are often undermined by learners' own subsequent statements, for example in the case of the above, which was actually preceded by: '*I feel like Arabic will get you nowhere*', or via responses to subsequent questioning regarding the likelihood of such a FS being achieved or pursued:

Student: ... If I move out to an Arab country or if you're going to move out to France or something, it's beneficial-- But I don't see the point of languages...

Researcher: Do you feel like you would move abroad at all?

Student: Um, no, I don't think so.

Such responses indicate the awareness of generic uses of language skills and the potential for their reiteration by learners who are completely disinterested in the goals they appear to voice. The ability to produce a description of a FS image is therefore clearly no guarantee of ownership or internalisation, making follow-up questions and more detailed measures of FS images imperative.

The second type of 'if' statement introduces a component of contingency into the IS equation and raises the question of plausibility.

Well, it depends if, how well I do, and how well I, how much I learn... But if I do do well, then yes, I see myself, cos I'm probably going to go back to Qatar because I love it there, so... and there, obviously, people's native language is Arabic, so... Yeah, I do see myself using it quite a lot in the future if I do well.

Multiple references (n=22) included descriptions of Arabic IS concepts as contingent on events in the nearer future, in particular, though not exclusively, attaining a good grade in their GCSE. The likelihood of a FS image becoming a reality was therefore dependent on more immediate factors, which, it was implied, were not entirely within learners' control. Whilst this contingency element could potentially be included in the 'plausibility' strand of the L2MSS IS, it has not been addressed specifically in the literature. This perception of a chain effect, of the realisation of a FS image as dependent on multiple stages, introduces a further level of complexity to the question of what constitutes a FS image in the first place. In assessing the likelihood of realising a FS image, learners need to assess the plausibility of succeeding in multiple pre-FS stages, each one of which is both a precursor to the ultimate FS image and a FS image in itself. In the quote provided, for example, doing well in Arabic is a FS, just as living in Qatar and speaking Arabic there is. The implication is, also, that if the more immediate FS is not realised, the longer-term FS is affected.

The third usage of 'if' statements can be interpreted to refer to less certain or internalised FS images. These are not the 'lip-service' statements of the first category (which do not involve any intended usage of Arabic in the future, since they were undermined by participants' own words), nor do they involve a contingency element. These are also, clearly, not emphatic statements of intended or desired future usage. Rather, these incorporate the element of possibility. Statements such as these often also include modifiers such as 'might' or 'maybe'.

Definitely, it might come into use, hopefully. And outside [...] if I decide to do any teaching on that side or anything like that, that might help, but as a wider career, I don't know... Maybe.

The words 'definitely', 'hopefully', 'might' and 'maybe' in this particular quotation capture a deterioration of certainty in relation to this IS concept, as if the expression of 'teaching' is first raised as a possible response to the FS image question, only for the learner to realise as they answer that this does not constitute an established component of their IS. The initial juxtaposing of 'definitely' and 'might' also appears to represent the awareness of possibility, followed by uncertainty as to whether or not this IS image is one the participant will actually pursue. Such statements of possible, but less integrated IS images stand in contrast to 'when' statements (n=9), which involve more categorical pronouncements of Ideal Selves.

Student: Yes, when I'm visiting other countries, insha Allah [God willing]. I will have to talk to people, yeah.

Researcher: Do you think about that a lot?

Student: Um... Yeah.

Researcher: Yeah? Have you visited Arab countries in the past?

Student: Yeah, I lived in Saudi Arabia for two years.

Here, there is no hint of dependency or uncertainty about intended future Arabic usage (other than the use of the phrase '*insha Allah*', 'God willing', which is employed formulaically in Muslim parlance when talking about the future such that it may be ignored here). What is significant in the above example is the link to past experience. Upon re-examination of the data, other evidence was found to suggest that certainty of future usage (in other words, more established, clearer IS concepts) was aided by concrete experiences with Arabic in participants' pasts. The use of 'when' has already been highlighted in the context of religious Arabic usage to imply a continuation of the present into the future. Here, there is an overlap in the sense that past experience of Arabic usage, most commonly either on holiday or with family, lends itself to adding plausibility to one's FS image. Using Arabic in the future becomes less hypothetical, based only on reports of potential usage, and more conceivable given the information and material supplied by the past for the creation of the FS image. The nature of the Past Self component will be discussed in Section 5.2.4.

4.2.7.11. Following the Thread: Interview and Survey Overlap

In line with the ‘Following the Thread’ approach employed earlier, learners’ self-expressed Arabic IS images detailed in the interviews are explored in the context of their self-concept item responses.

The 95 interview participants were divided into groups in SPSS based on their references to the nine categories of future self-images. The three most frequent categories were *Travel*, *Religion*, and *Job*, (see Table 4.11) which, incidentally, could be conceived as overlapping with Dialect, Classical, and MSA respectively. Independent samples t-tests were then run to examine whether references to these FS images corresponded with significantly different scores for any of the three varieties included in the survey scales.

Table 4.11. Frequency of Arabic Future Self References	
Arabic FS Category	Frequency
Travel	51
Family	9
Religion	30
Job	25
Skill	11
Other	2
Language Study	1
GCSE	5
Future Study	17

Given the number of t-tests run, the recommended Bonferroni adjustments were made to the p-value within each set of tests ($p < .006$). When comparing learners who had mentioned *Religion* and those who had not, no significant differences in survey scale items emerged at either timepoint. Although this result does not provide conclusive proof given the limited sample sizes, this suggests that explicit reference to religion in interviews does not correspond to a higher CS, IS or OTS measures for Classical Arabic. This seems to indicate that the absence of an expressed Classical Arabic oriented FS image in the interviews does not mean its absence from learners’ goals and CS concept, given its ubiquity in their lives. This, however, does not explain why the mention of *Job*

did not align with higher scores for any specific variety. The actual explanation for this, however, may simply be attributed to learners being unaware of which type of Arabic they may require for the purposes of a job, or that they wish to incorporate multiple varieties into their future self-concepts.

T-tests for *Travel*, on the other hand, resulted in a notable area of significant difference. Participants who mentioned Travel as a FS image in their interviews scored significantly higher in their CS scores for all varieties at T2 (MSA: $t(51)=-4.054, p=.000$; Classical: $t(51)=-2.981, p=.004$; Dialect: $t(51)=-3.575, p=.001$). In other words, learners who saw themselves as likely to travel in the future had stronger CS concepts at the end of the academic year. The direction of influence between mentioning travel and stronger CS concepts is not ascertainable from these results. It may be that learners who feel more confident in Arabic see themselves as more likely to travel. Alternatively, it may be that wanting to travel has made them exert more effort to improve their skills and therefore their CS concepts. Again, however, this did not translate into stronger IS concepts in the survey data for these students. It is plausible that learners are aware of the possibility of travelling to Arab countries without the necessity of using Arabic in these contexts. The following quote, in particular, captures this awareness:

... Only if I go on holiday or something like that... That's only if- if I think about it I just use iTranslate, or Google Translate, I just use that.

Here, the use of translation technology is cited to explain the reason for a lack of future use of Arabic even in the context of desired (and past) travel abroad. The absence of a relationship between the expression of travel to Arab countries as part of the FS image and stronger IS scores may also be explained by the occurrence of 'if' statements in this category. Of the 49 references to Travel in the interviews, 13 of these were 'if' statements (described in Section 4.2.7.10). The fact that 27% of references to travel were hypothetical, contingent, or uncertain offers a potential explanation for why learners in the Travel category were not observed to have stronger IS concepts.

4.2.8. Ought-to Self

In order to determine the nature of participants' OTS concepts, two questions were asked: Whether participants felt any pressure to learn Arabic, and whether there would be any consequences of not doing well in the subject. Learners were not asked directly whether or not they felt they 'ought' to learn Arabic, largely because this phrasing was considered less transparent for teenage learners, and the identification of sources of pressure and potential consequences tapped into the underlying components of OTS concepts as outlined in the theoretical framework.

It was evident from the analysis of responses to the first of these questions that learners interpreted 'pressure' to mean two things: a sense of obligation regarding the learning of Arabic (whether self- or externally imposed); and the stress associated with studying for the Arabic GCSE. The latter contributed to learners' CS concepts since it provided details as to how learners perceived and felt about their current language learning progress. The former was used to identify three sources of a sense of obligation or necessity to learn Arabic: *Parents*, *School*, and *Self*, the last of which referred to learners' own self-imposed sense of obligation.

As far as consequences were concerned, there were large areas of overlap with the former categories, such as the fear of disappointing parents or teachers. However, a major consequence identified by learners of lack of success in their Arabic GCSE was the negative impact on their general FS concepts. In this sense, it appears that there are two strands to participants' OTS concepts—areas where a sense of obligation contributes an element or intensity to a FS image, and those where the emphasis is on avoidance of a particular FS concept component. This second strand, it could be argued, aligns more with Dörnyei's 'Feared Self', which he makes brief mention of in his original description of the L2MSS (Dörnyei, 2009). The following discussion on themes emerging from an examination of the data for insights into the OTS is organised according to the three sources identified, each of which contain elements of both the obligation and consequences strands.

4.2.8.1. Self

4.2.8.1.1. *Future Plans*

Despite the numerous descriptions of potential and plausible Arabic IS images, few learners made an explicit link between not doing well in their Arabic GCSE and the impact on these Arabic-specific FS images. Some students did mention that they would continue to pursue the study of Arabic, either by retaking the exam, or through other means outside of school. This means that achieving these IS images would still be possible and offers a potential explanation for the absence of a link between GCSE and future success. What learners did make mention of in relation to consequences of failure was the impact on their non-Arabic specific FS images.

Student: The world is pressuring me. Because like uni, if you don't get good grades, then you might not really do well in life. Like GCSEs are really important. As we get closer to GCSEs, it has kind of become more realistic. No one really pressured me but my parents, they do want me to try. They don't care about Arabic as much as other subjects, the main ones like Maths, English, science... they would nag me. They would be happy if I did do well in Arabic, but they don't really pressure me. [...] it's kind of overwhelming, because a lot of people say that if I don't do well in the exams, we're kind of done.

Researcher: Who says that?

Student: Everyone.

This overlaps clearly with the theme of contingency (see Section 4.2.7.10), in the sense that a FS concept is the result of the satisfaction of multiple cumulative requirements. Arguably, the above quote represents the flip-side of this line of thinking; the domino effect of failure on a FS image. Regardless of whether or not that image involves Arabic, the language GCSE has become a steppingstone or potential sticking point on the path to a FS concept.

Not all students who referred to the impact of their GCSE grades on their FS images in a theoretical sense, followed this line of thinking all the way to future jobs, many of them talked of the immediate impact on college choices. It could be argued, however, that implicit in this statement is the line of FS 'dominoes'. Interestingly, as the participant in

the above quote mentions, there appears to be little, if any, imposed expectation on them to do well specifically in Arabic. However, as this subject does constitute one of her grades, this may be the source of this sense of compulsion.

In terms of the implications on future research, such references, which include the conflation of the Arabic language and the GCSE subject, add further evidence for the need to distinguish between these two constructs in the context of school learners' current and future self-concepts.

4.2.8.1.2. *Waste Avoidance*

Many learners described their own self-imposed sense of pressure in the context of learning Arabic. One notable source of this sense of obligation was the necessity of 'waste avoidance'.

I'd feel upset, because like he [classmate] was saying, five years for, like, nothing.

I guess I'd feel disappointed in myself because it's five years of work that goes down the drain and as he said as well, in terms of getting into better schools for A Levels and those kinds of things.

The majority of students in the sample had studied Arabic for at least 3 years, most of them having joined their respective schools in KS3. One of the reasons they feel they should do well in Arabic (as a short-term FS) is the amount of time they have been obliged to invest in learning the language in school, in other words, failure would be a waste of their time and effort. This falls under the source banner of *Self* since it is neither their parents, nor the school imposing this particular view, but the students themselves. This personal sense of obligation—to avoid waste—adds credence to the notion that the OTS does not necessarily align solely with external sources of motivation. Rather, as Markus and Nurius (1986) outline in their original theories, impetus for the OTS can come from within as well as without. Future instruments which measure the intricacies of the OTS, and indeed its validity, need to take into account this particular source of obligation.

4.2.8.1.3. *Disappointment Avoidance*

Another component of the OTS which falls under the banner of consequences, is the fear of or desire to avoid disappointment. The possible consequence of regret, or disappointment in oneself, features prominently in learners' responses to this question.

I would be upset with myself, because I knew I was trying really hard for that subject. And obviously I failed, like, you can't be happy about that. Like, I would just be so sad.

I think I'm just drawn to the language, so, like... It will hurt a lot if I fail in that language.

Interviewees do not mention explicitly that they are studying Arabic in order to avoid a self that is afflicted with such feelings. However, the awareness of the potential for such FS images, it could be argued, is an implicit OTS concept, if not a Feared Self which works in tandem with the IS.

A particularly illustrative example of such a FS is provided by one of the students who had already sat her GCSE that year and who feared she had not done well:

Student: I think I'll get really really, really upset [if I don't do well]. I didn't come in for a week after the exam. To school.

Researcher: You didn't come in for a week after the exam?

Student: [...] Cos of the exam. Cos I was just really really upset. And then the Headteacher had to phone my mum cos- (laughs) And tell her that I had to come in- Cos I put in so much effort- And then when the writing came, it was just completely different. I wasn't prepared for it. Like, I was prepared... In other topics. And I worked really, really hard- I remember I didn't, like, go to sleep for, like, one week. And then- I feel like my, my whole hard work, everything, was wasted. So, it was kind of sad...

4.2.8.2. *Parents*

Two codes, *Consequences: Parents* and *OTS - Pressure: Parents*, captured the role of parents in contributing to learners' OTS concepts. As suggested by the names of the codes, these contributions took two major forms: parental expectations upon students

to study and/or perform well in Arabic, and the potential for parental disappointment if learners' do not succeed in their Arabic GCSE.

4.2.8.2.1. *Parental Expectations*

A number of students who discussed their parents' interactions with their Arabic learning described their parents as being supportive and encouraging of learners studying Arabic, and not exerting pressure on them to do well or continue with the language.

I don't feel really pressured, but I do feel that my parents would- if I do achieve that then they would be happy [...] But... they don't really pressure me to do it. Like... after the GCSE, they don't mind if I don't take Arabic. Yeah, so, then I don't really feel that much pressure.

Like my parents don't pressure me at all. They just want me to try my best.

Other students, however, communicated the fact that their parents did want or expect them to a) study Arabic in the first place, and b) to do well in the subject. This aligns with a division of motivation into *initial* and *continued* strands, which often but do not always overlap, adding credence to the need to separate the two in the L2MSS framework. It was not often clear where the distinction between these two motivational domains lay; in some instances parental expectations were expressed in reference to the initial option of studying Arabic. Learners in one class even spoke of their parents coming into school to insist that their children study Arabic rather than French, given its religious significance. This impetus is also indirectly implied by parents' choice to send their children to Islamic schools in the first place, given the integration of Islamic Studies subjects into the curriculum.

As the section below on the role of religion in the OTS makes clear, learners do not themselves experience direct religious Ought-to Selves. Participants in this sample did not communicate their own sense of obligation to learn Arabic for religious purposes. Rather, they expressed a desire to do so, presenting it as a component of their IS images. This apparent absence of religious Ought-to Selves was rather striking. However, an analysis of learners' comments on their parents' expectations makes it clear that religion

features in their OTS in an indirect way. In other words, expressions of explicit parental expectations were commonly tied to religion, in the sense that parents wanted their children to learn Arabic in order for these learners to apply it to their religious practice.

I mean it's not completely forced, obviously, but it's definitely pressured because... Obviously [...] because we're all Muslim and parents want us to learn Arabic so that we can, we can, you know, use it to just to understand the Quran more, to be able to interact with people, in like, the Arabic world, it was the language of the Prophet... Um, yeah, so I guess we're a little bit... A bit pressured into learning Arabic. More like, I guess more like encouraged, but yeah. It's kind of the same thing.

What is significant about this particular quote, which alludes to the difficulty in identifying Ought-to Selves, is the distinction between, firstly, *force* and *pressure*, and subsequently, the equating of *pressure* and *encouragement*. The fact that learners are not forced by their parents to learn Arabic implies a lack of strict obligation; however, the pressure component does imply an expectation, or at least awareness of an external imposed desire. It is possible that this particular learner has chosen to interpret this expectation as positive and therefore framed it as encouragement. Any such reframing, however, does not represent an internalisation of the desire into an IS image, meaning, potentially, that it is possible for OTS images to be perceived as favourable, and indeed helpful, while still not necessarily being of one's own generation.

Learners from heritage backgrounds also mentioned their parents' desire for them to be able to speak Arabic for communicative purposes:

Obviously because my mum speaks Arabic as well- And English. So she was always, like, 'If you don't want to learn Arabic, then don't learn if you don't want to. But it's better for you- 'So if you go back home you're not embarrassed,' because everyone's talking to me in Arabic and you're just sitting there, like, don't understand what you're saying. It'll just be embarrassing for you, but if you don't want to learn it then you don't have to, because it's not like we're gonna use Arabic every single day of your life. Of course, it'll come in handy when you're, like, mingling with people...

This often overlapped with learners' IS descriptions; however, it is not clear whether the overlapping of learners' IS and OTS provides additional motivation for students. Participants spoke of their parents' general grade expectations. This was either

presented in the context of sibling comparison, in instances where participants' older siblings had studied the language and achieved good grades in previous years, or a desire for these participants to simply attain high grades. There was an apparent split as to whether parental emphasis was on getting good grades in all subjects, including Arabic, or whether Arabic was seen as a priority subject. In terms of the latter, some learners mentioned the 'prestige' and 'benefits' of Arabic in comparison to other subjects as perceived by their parents. It appears that, to some parents, the value of Arabic was far more transparent than that of other subjects:

... When you tell... Islamic parents that, oh, you've got a Arabic GCSE... They're more focused on that than about, you know, History, about you know, dead people and stuff like that.

4.2.8.2.2.

Disappointment Avoidance

The second dimension of learners' Ought-to Selves which involved their parents related to the potential for parental disappointment in response to lack of success in Arabic. This was a common theme in learners' responses to the question regarding consequences of not doing well in Arabic. 12 students made reference to parental disappointment in their answers, though it should be mentioned that many students also expressed that their parents would not be disappointed if they didn't do well. Rather than a binary division between parents who would and wouldn't be disappointed, there seemed rather to be a scale of disappointment, from 'a bit' to 'quite' to 'really disappointed' and even 'angry' and 'upset'.

Those who did express the potential for parental disappointment were split, again, as to whether this reaction would be exclusive to Arabic, or whether a similar response would be elicited following lack of success in any subject. Those who mentioned the latter and elaborated on the reasoning behind this response predominantly referenced religion.

Whilst the theme of disappointment avoidance was clearly prominent in the data, the question remains as to the precise role of avoidance of potential FS in the L2MSS. This avoidance factor could theoretically constitute either part of an individual's OTS, or a

Feared Self, the flip-side of the IS—the existence of which, according to Dörnyei's theory, should inspire stronger motivation. The prevalence of this theme, and its role in the L2MSS, if it has any, remains shrouded, necessitating further research.

4.2.8.3. School

Though not explicitly expressed as an OTS concept, the fact that many of the schools in the sample impose Arabic as a compulsory subject at GCSE, or offer an imposed but restricted choice (usually between two options) for languages was mentioned by learners in response to the question regarding pressure. The implication of the imposition of Arabic by the school and the aforementioned choice of parents to send their children to Islamic Schools is that knowledge of Arabic should constitute a part of students' identity and skill repertoire.

These references were coded as *OTS—Force* and *OTS—Pressure: School*. As was the case with the theme of 'pressure' overall, it was necessary to distinguish between the stress learners were feeling when studying Arabic and the sense of pressure they felt to study or continue to study the language.

When referring to the initial 'choice' to study Arabic at GCSE, many students mentioned feeling or being '*forced*', either through explicit use of this word, reference to Arabic as '*compulsory*', or phrases such as '*we didn't have a choice*'; '*we had to do it*'; and '*they just gave me Arabic*.'

Regardless of whether or not learners had chosen the language, reference to the obligation to continue with study was also made, indicating that once enrolled into the Arabic GCSE, students were not able to opt out:

It's just the teachers putting pressure on us and I've spoken to a lot of teachers about not doing Arabic but they said, 'Now you've started the course, you have to finish it now.'

References to the feeling of being forced to study Arabic were often made alongside expressions of despondency (see Section 4.2.6.2.1.1). One student went further to describe the effects of being made to study Arabic without any element of choice:

Because we don't choose our subjects. You know, they're chosen for us by the teachers. So... we don't have that privilege to, you know, be passionate about certain stuff [...] In a state school you get to choose something that you know that you love... Cos when you ultimately love something, you do better in it.... Cos we don't really love any subjects we do- It kinda puts us in a disadvantage. [...] Because I feel like when you love something, it's from your own will. It shouldn't be put on top of you- Like, you should love it from your own- You know, decision that all the subjects I really enjoy- It shouldn't be, like, "You do this subject, you have to do it, you can't drop it out..." So you're, like, you're more, like, forced to do it, in a way? So you're just, like, "Ugh, I just have this subject [inaudible]- I don't want to do this..." Like, certain days you look forward to, certain days you don't look forward to because of the subjects.

Feelings of despondency are apparent in this quote, alluded to by the students' feelings towards days on which they have Arabic lessons. However, what is most profound is their reference to the impact of a lack of choice, which they describe as a 'disadvantage'. The imposition of the subject as compulsory, coupled with the fact that this learner (and seemingly their classmates, as suggested by the use of 'we') does not 'love' the subject, negatively impact upon their affective interactions with the subject. Though participants do not mention it themselves, negative emotions can affect learners' achievements in a subject (Dewaele, Witney, Saito, et al., 2018; Pekrun et al., 2017). Moreover, the mention of deprivation of the 'privilege' of choice, perceived by this student as a feature of state schools, could be interpreted as a sense of disempowerment and raises worrying questions. What are the ethical implications of schools choosing to impose a subject, which: a) is presented as an integral part of the religion, and therefore learners' religious identities; b) learners not only perceive as 'forced' but also do not enjoy; and c) as a result of which they are made to feel despondent, and which they are not even allowed to opt out of?

4.2.8.4. Religion: The Missing Ought-to Self

One striking finding in the context of this study is the absence of direct references to

religion in the context of learners' Ought-to Selves. Given the fact that a number of rituals in the Muslim faith are performed in the Arabic language (Rosowsky, 2006; 2008), and that such rituals are often referred to as 'obligations' or *furūd*, in Islamic terminology, religion was hypothesised to feature strongly in discussions of learners' OTS concepts. This was not the case. It did appear, however, that learners experienced some of this sense of obligation indirectly. As referenced above, learners reported feeling pressure and/or a sense of expectation to learn Arabic via their parents or school. In both cases, links were made between these sources of obligation and religion; parents expect their children to learn Arabic in order to be able to read the Quran, and schools have made it compulsory given their Islamic ethos and curriculum. These findings are discussed further in Section 5.2.3.2.

4.2.8.5. Gender

As with the CS and the IS, no significant differences were observed between genders at either timepoint in the realms of the OTS. An analysis of the interview data raises two interesting findings. The frequencies of references to different OTS codes, organized by gender, are included in Figure 4.15. One is that a higher percentage of male students referenced threats to future plans as a result of not doing well in Arabic, in particular the impact on their Sixth Form choices: *'I'll probably get rejected by every Sixth Form...'*. 4 out of 5 of these came from Class 5, which suggests a school effect, though this cannot be corroborated by the data directly.

Another suggestion of school effect, however, is the higher level of explicit *School Pressure* and *Compulsory* category amongst female students, indicating, potentially, that certain schools, in this case the girls', apply more noticeable pressure on their students. There does not appear to be a difference in the amount of parental pressure or consequences learners are subject to, nor in the pressure they put on themselves.

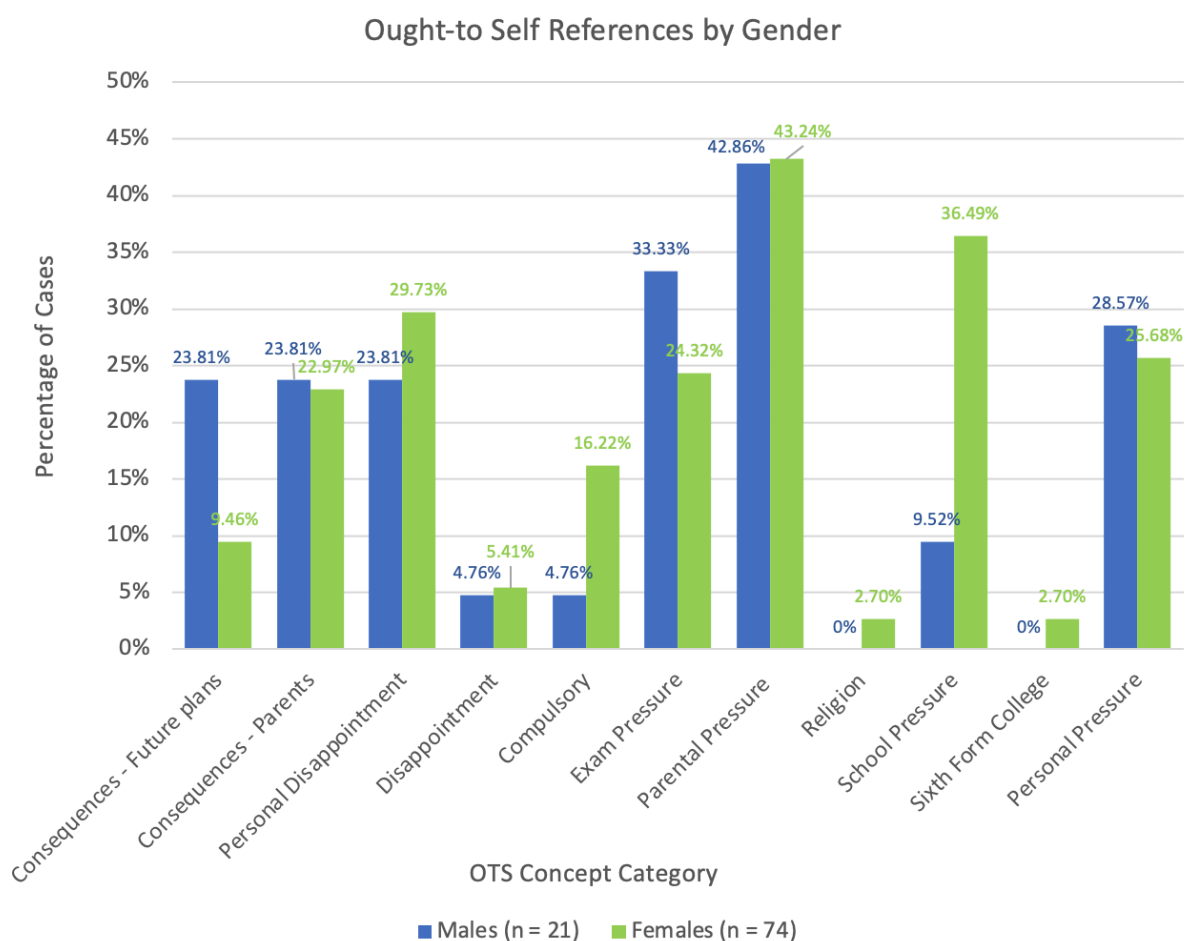


Figure 4.15. Bar graph for references to OTS codes by gender

4.2.8.6. Heritage Learners

At T1, NHLs' OTS scores were higher than those of HLs, but this was reversed at T2. Given that the interviews were conducted in the same time period as the T2 surveys, it may be assumed that the interviews would corroborate this finding (bearing in mind the change and limitations in sample sizes).

At first glance, the disparity in scores between HLs and NHLs, represented in Figure 4.16, for Personal Disappointment, Parental Pressure, and Personal Pressure, for which HLs scored almost double appear to suggest the sources of these learners' OTS scores more strongly derive from their own self-imposed pressure and consequences, along with pressure from their parents. This is not to say that NHLs do not experience parental pressure. In fact, 25% of NHL participants mentioned the possibility of parental consequences if they did not do well, which largely revolved around

disappointment. To return to the earlier distinction between obligation and consequences, it appears that HLs experience a stronger sense of obligation from their parents towards learning Arabic, whereas NHLs feel a stronger need to do well in order to avoid potential consequences from them. Another interesting area of distinction are the higher School Pressure scores for NHLs, almost double that of HLs. Though the scope of this study does not allow for an examination of this, one hypothesis is that learners who do not exhibit strong IS images are more likely to feel more pressure from school, particularly where the subject is compulsory, since they do not possess their own motivation through which to frame the efforts they make in the classroom in order to do well.

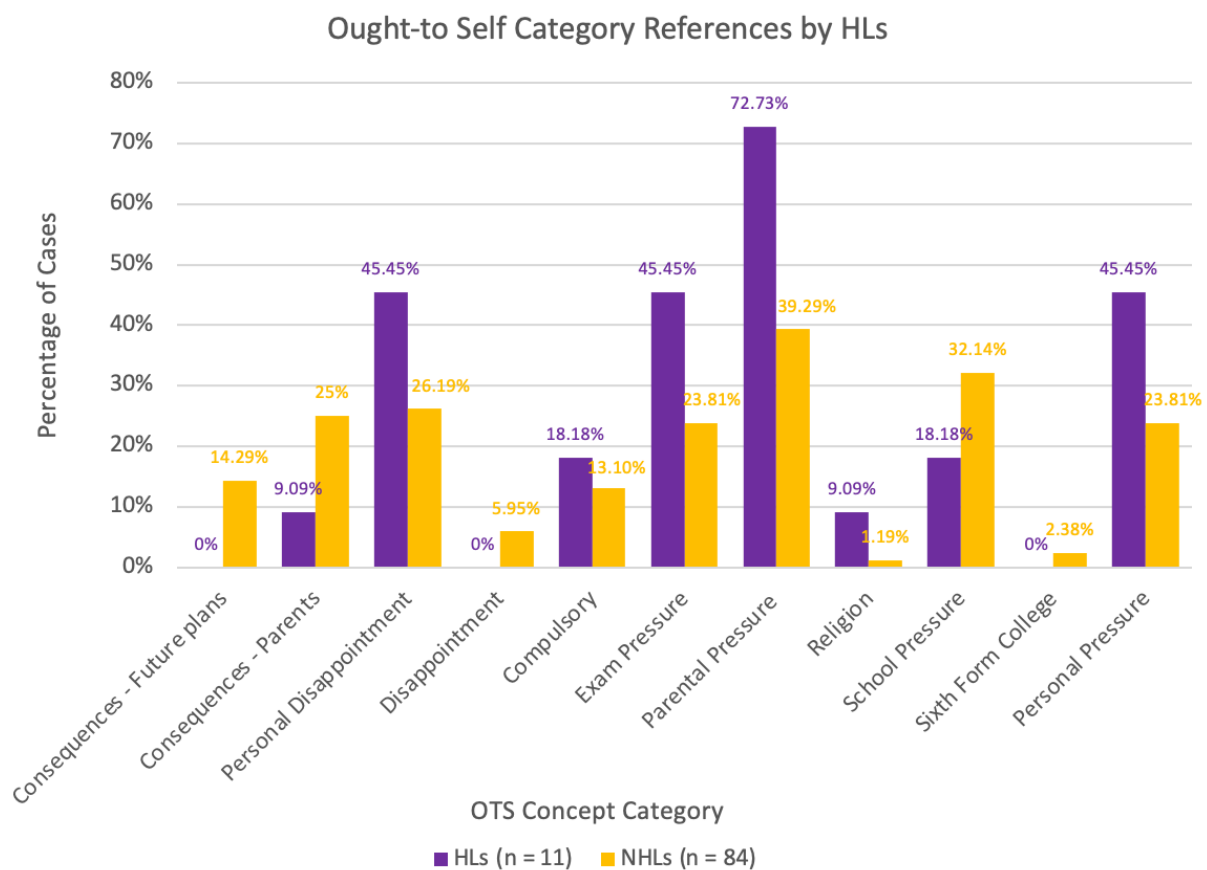


Figure 4.16. Bar graph for references to OTS codes by heritage background

4.2.9. Past Self

Literature on the L2MSS identifies the current and future self-concepts and the discrepancy between them as the main area for focus when attempting to capture learners' motivation. The interview data collected from the sample in this study, however, highlights a major component absent from this framework: the Past Self.

The Past Self incorporates references to both (a) individuals' confidence, ability, and learning trajectories, coded in the interview data as *Past: Arabic Learning*, as well as (b) experiences with Arabic outside of the educational domain, labelled *Past: Personal Experience*. These two constituents of the Past Self, and their relationship to Current and Future self-concepts, and their integral nature in the context of motivation are explained below.

4.2.9.1. Past: Arabic Learning

When asked to discuss their Arabic learning experience, students commonly made reference to their previous years of study, including the amount and the nature of their Arabic learning prior to their current year. As mentioned in the section on the CS (Section 4.2.6), these past experiences appear to be integral in the formation of learners' CS concepts because they provide a frame of reference against which to judge their current levels of confidence and ability in the subject. CS concepts do not exist in isolation of their temporal contexts: learners' self-assessments are relative to their past learning experiences.

... At the beginning, like Year 7 I was really bad at Arabic. And then from Year 7 to now, I got better. But I'm not that good.

... In the early years of secondary school, when I first started Arabic... it wasn't very difficult for me, cos I've been studying Arabic from a really young age. So when I came in, I had good grades in Arabic and everything. But then as soon as GCSEs started, it was really difficult in the beginning of this year, because there's a really big gap between what we were learning in the earlier years...

4.2.9.2. Following the Thread: Links between Past Learning and Self-Concepts

To explore whether or not the past learning component contributed to learners' self-concepts, the correlations between the number of years learners had studied Arabic at school and their Current or Ideal Selves were examined. Significant correlation coefficients are provided in Table 4.12. Though very weak correlations were observed between Years of Arabic at School and CS concepts for all varieties, and between Dialect IS, and Classical OTS at T1, no correlations were apparent between years of Arabic at school and any of the variables at T2. This suggests that past learning experience has little bearing on the construction or development of learners' Ideal and Current Selves.

Table 4.12. Significant correlations between Arabic at school and Self-concepts					
	D CS Sum Score	D IS Sum Score	CI CS Sum Score	CI OTS Sum Score	MSA CS Sum Score
Years of Arabic at School	.132*	.121*	.188**	.164**	.136*
Sig. (2-tailed)	0.032	0.049	0.002	0.008	0.029
N	262	264	264	263	260

Since comprehensive methods of learners' past self experiences were not included in this study, an analysis of the links between Past, Current, and Future self-concepts is limited. However, an examination of the data suggests no coherent pattern between Past and Current selves in the realms of confidence or proficiency. Examples can be found of learners having had high levels of confidence in the past and lower ones later, higher ones now, or similar levels of confidence. The same can be said for low past levels of confidence.

One trend which was prominent across the sample relates to past and current levels of enjoyment. As explained in relation to the Current Self, affective components, though largely overlooked by the L2MSS, appear to play an important role in self-concept construction (see Section 4.2.6.2). Many learners spoke of enjoying their Arabic learning more during KS3 years, in comparison to the difficulty they now faced when studying for the GCSE. It seems that the contrast between the Key Stages accentuates the

difficulty currently faced (exacerbated by the change in specification; see Ramezanzadeh (in preparation)), which often leads learners to lower levels of confidence, lower self-assessments of their abilities, and potentially, despondency:

I've been doing Arabic since Year 1. So when I first came to this school, Arabic was, –it was okay for me. But then, like they've all said, as I got more into it, especially in Year 10, it became more harder for me, because the specification changed and just the thought of how much we have to know for the exam-- It just made me stress out – and because I have other subjects, I decided to, like, not really leave Arabic, but, put my focus on other subjects, because I feel not confident in Arabic, so I feel like there's no point in doing it, because I don't feel like I will do well in it, cos-- it just stresses me out knowing that we have to do so much and in so little time.

4.2.9.3. Past: Personal Experience

Another feature of learners' Past Self descriptions relates to the relationship between past experiences with Arabic and the construction and nature of their FS concepts. Whilst the evidence for this is not conclusive, and future research into this area is necessary, there appears to be a link between learners' past experience with Arabic, in the realm of the personal, and a) the availability of their FS images, and b) the nature of them.

Examples of *Past: Personal Experience* include, predominantly, HL experiences communicating with their Arabic-speaking families/friends:

I have family abroad, and I have a grandma who... speaks more Arabic than other languages so I want to be able to communicate with her and I feel like Arabic would help me with that.

NHLs communicating with family/friends in Arabic-speaking countries:

... Now I have family that speak Arabic and stuff. Though my cousin, she speaks Arabic, she speaks English as well but, like, it's not as fluent. Yeah, so she really wants me to study Arabic. She's like... "You have to study Arabic." She'll try teaching me and I'm just, like, 'Yeah... Okay...

Witnessing communication/communicating directly with Arabic speakers on holiday:

... When I went [on] 'umrah [pilgrimage] with my family, [in] the year 2016,

my sister came with the school and she knew Arabic as well and she could ask where the toilet was and where's this and that. So yeah, I think it's quite useful in visit[ing] Arab countries.

In one case, a learner also referred to their work experience in which they were able to utilise their mother tongue to interpret for a patient in a hospital setting, which they then realised they could also do with Arabic if they were to learn it.

Theoretically, the link between past and future selves is a logical one. FS images are not constructed out of thin air: the material for a FS image must be supplied before it can be internalised and elaborated upon. Whilst this may come from hearing about potential future uses for Arabic, which does not necessarily appear to be effective in generating genuine FS images amongst participants, it may also come from direct experience with Arabic in the real world. Traditionally, HLs display much higher levels of motivation than their NHL counterparts. However, the reason for this has not been explained in depth. Why would a HL's FS image be so much more available and vivid than that of their NHL counterpart? One likely answer is that learners are able to visualise themselves more clearly using Arabic in the future because the context in which they can do so is already being/has been experienced. To a lesser degree, it seems, this also applies to NHLs. Several students mentioned that they had been to Arabic-speaking countries on holiday in the past. Some of these students recounted their experiences trying to speak Arabic during these episodes. An examination of the Arabic FS images of these learners suggested that they enjoyed a more established FS image. This is expected since they, like HLs though to a lesser degree, have been immersed, even if briefly, in a context in which Arabic is actually used. This serves to move their FS out of the realm of the purely hypothetical and into the realm of the plausible. A similar situation is apparent in learners' references to the religious application of Arabic. This is clearly a well-established component of their past lives and constitutes an integral part of their FS images, even when they do not explicitly mention it as such.

In order to investigate this further, interview participants were categorised into those with Past Experience with Arabic (PE) and those without (NPE). Frequencies of references to past selves in the interviews were then cross-tabulated with categories of Future Self images (see Table 4.13). What is perhaps unsurprising in this table is that all

of those who referred to Family as constituting part of their Future Self image were those in the Past Experience (PE) category.

Interestingly, though also perhaps predictable, is that there is an equal split between PE and NPE categories in mentioning Religion as a constituent of their FS image, indicating the ubiquity of religion-related goals amongst learners regardless of their exposure to other types of Arabic.

Table 4.13. References to AFS codes categorised by Past Experience								
Nodes	% of Cases per Category						% of Total Cases per Node	
	Past Experience (n=39)		No Past Experience (n=56)		Total (N=95)		Past Experience (n=39)	No Past Experience (n=56)
Family	12	31%	0	0%	12	13%	100%	0%
General Communication	6	15%	7	13%	13	14%	53%	47%
Job	13	33%	13	23%	26	27%	59%	41%
Just GCSE	6	15%	4	7%	10	11%	54%	46%
No Arabic	4	10%	21	38%	25	26%	11%	89%
Other	0	0%	4	7%	4	4%	0%	100%
Religion	15	38%	15	27%	30	32%	43%	57%
Skill	10	26%	11	20%	21	22%	44%	56%
Travel	24	62%	19	34%	43	45%	69%	31%
<i>If' statements</i>	13	33%	23	41%	36	38%	46%	54%
<i>When' statements</i>	8	21%	3	5%	11	12%	77%	23%

What confirms the premise that Past Experiences are related to Future self-concepts, in the sense that they promote them or supply the necessary experience upon which to base them, is the finding that 89% of learners who expressed an entire lack of Arabic-related FS concept had no past experience with Arabic. Further evidence for this proposal is provided by the frequencies of 'if' and 'when' statements within each of the categories. 54% of those in the NPE category expressed 'if' statements (e.g., '*if you're going to travel*'), in comparison to 77% of those in the PE who expressed more definitive 'when' statements when describing their Arabic-related FS images (e.g., '*when I'm visiting other countries*').

To examine whether or not a relationship between the Past Self and the CS exists, a further set of nodes, concerning *Feelings*, were cross-tabulated with the Past Experience categories. Feelings, it was identified in the qualitative analysis in Section 4.2.6.2, constitute a theoretically integral, but overlooked, part of learners' Current Selves.

The distinctions in the frequencies for these references, provided in Table 4.14, were even more striking. Higher levels of overall negative feelings towards their Arabic learning experience were apparent amongst the NPE category, 47% of whom expressed feelings of despondency with their Arabic learning experience, constituting 82% of the overall occurrences of cases in which this theme was mentioned. This is in comparison to only 16% of learners in the PE category who referenced feelings of despondency. In contrast, higher levels of enjoyment and overall positive sentiments were observed amongst PE participants, 49% of whom expressed positive feelings compared to only 25% of NPEs.

Table 4.14. References to Feelings codes categorised by Past Experience							
Nodes	Past Experience (n=39)		No Past Experience (n=56)		Total (n=95)	Past Experience (n=39)	No Past Experience (n=56)
Apathy	2	5%	8	14%	11%	20%	80%
Despondency	6	16%	27	47%	35%	18%	82%
Dislike	3	8%	2	4%	5%	60%	40%
Enjoyment	11	29%	6	11%	18%	65%	35%
Negative	7	18%	19	33%	27%	27%	73%
Positive	19	49%	14	25%	35%	58%	42%

These findings suggest a link between past experiences with Arabic and the affective components of learners' Current Selves. While causation cannot be implied from this particular set of data, these findings offer strong evidence for the further investigation of this subject with larger sample sizes and further longitudinal designs.

4.2.9.4. Following the Thread: Links between Past Experience and Self-Concepts

In order to follow this thread back through the quantitative data, learners who had participated in both the interviews and surveys at both timepoints were identified and isolated in the quantitative data. In total, this constituted 81 participants for T1 data and 53 at T2. Based on their interview responses, participants in this dataset were divided into PE and NPE categories in order to compare whether or not the differences in Future and CS concepts detected in the interview data were apparent in the equivalent quantitative constructs.

4.2.9.4.1. *Normality tests*

The z-scores for the skewness and kurtosis statistics for the motivational self-concepts and proficiency score data were calculated for participants in both groups in order to analyse whether or not the data was normally distributed and therefore appropriate for parametric testing. The cut off score for suitability was set at ± 1.96 since the size of each subgroup was < 50 (Kim, 2013). Given that nine of the items exhibited skewness scores above the 1.96 cut-off point, non-parametric tests, namely Independent Samples Mann-Whitney U tests, were used in order to compare the means of these two groups.

4.2.9.4.2. *Mean Comparisons*

11 tests of T1 variables were run, with the equivalent at T2, necessitating a Bonferroni corrected p-value of .005 in order for a finding to be declared significant. In line with these stipulations, significant differences were observed between the PE and NPE groups in a number of variables, the results of which are presented in Table 4.15.

Table 4.15. Mean Comparisons for Self-concepts by Past Experience (Significant results only)						
	Variable	N	Mann-Whitney U	Standardised Test Stat	Sig.	Medians
T1	CS - MSA	81	1097.50	2.936	.003	NPE (-0.45) PE (0.28)
	CS - Classical	81	1096.50	2.927	.003	NPE (-0.22) PE (0.56)
	CS - Dialect	81	1076.50	2.735	.006	NPE (-0.59) PE (0.36)
	Overall CS	81	1104.50	3.004	.003	NPE (-0.44) PE (0.35)
T2	CS - MSA	53	528.00	3.150	.002	NPE (-0.77) PE (0.37)
	CS - Classical	53	509.00	2.812	.005	NPE (-0.70) PE (0.63)
	Overall CS	53	517.00	2.955	.003	NPE (-0.51) PE (0.42)
	IS - MSA	53	556.00	3.647	.000	NPE (-1.20) PE (0.50)
	Y/N Proficiency Scores	60	658.00	3.114	.002	NPE (-0.69) PE (0.14)

What is striking here is that at T1, the existence of past experience translates into higher scores, across all varieties and overall, for learners' CS concepts. However, past experience did not influence participants' Ideal or Ought-to Selves, nor their proficiency scores. By T2, these differences expand to incorporate MSA IS and Y/N Proficiency scores. This seems to suggest that between these two timepoints the existence of past experience (or perhaps the gaining of experience during that year, since the interviews were conducted at T2) contributes to the development of a stronger MSA IS and higher scores in the vocabulary-oriented test. It is possible that learners' experience with Arabic during that year contributed to their acquisition of vocabulary, which led to higher Y/N scores at T2. An alternative explanation is that the presence of Past Experience provided contexts for the usage of certain high frequency vocabulary items (which were measured in the Y/N test) and this facilitated or promoted the process of acquisition and subsequent retention.

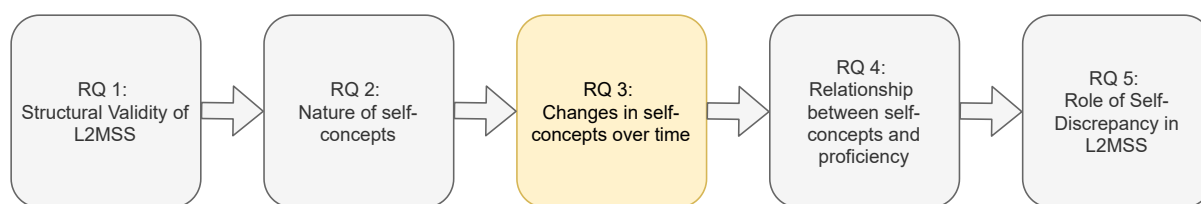
The translation of past experience with Arabic (usually gained from travelling abroad or contact with Arabic-speaking family members) into higher IS scores for MSA is noteworthy. One would expect experience communicating with NSs to promote a

learner's Dialect IS, however, this was not the case. Why did significant differences emerge predominantly in the realm of MSA, particularly when this was expressed in the surveys in terms of the ability to read the news, books, and magazines? This variety emerged consistently as the lowest in terms of Current and IS scores, and was the least frequently referenced in the interviews themselves. Possible explanations for this include a lack of ability to distinguish between colloquial and standard varieties and the subsequent belief that an ability to read Arabic will facilitate future goals, such as travel and general communication. There is also the possibility that experience with Arabic does in fact promote the desire to read Arabic, though this is not expressed by learners in the interviews. Given the overlap between the MSA variety as described in the surveys, and the content of the Arabic GCSE, this could indicate a stronger desire to pass their GCSE, which involves a large reading content. Perhaps learners believe that GCSE level knowledge would have facilitated their past experiences in Arab countries, for example. Indeed, some learners hint at this potential explanation:

I just went to Saudi Arabia in April, in the Easter holidays. And even doing it for about a year and a half, back then, I didn't really understand too much. I could make, like, a very simple conversation. That's basically it. So if I want to carry on so I can actually speak to someone in the Arab world, or maybe, I don't know, live there.

Whilst further studies in this regard are clearly required, these tentative findings appear to suggest the invaluable role of past experiences in promoting learners' IS concepts in the realm that is most aligned with the actual content of the GCSE. Teachers and curriculum developers would be well advised to work on facilitating the acquisition of authentic experiences with Arabic, so as to harness the apparent effects of this on learners' motivations.

The incorporation of the Past Self into the L2MSS, thus, appears to be essential, since it contributes to the construction of both the CS, by way of providing a frame of reference and by acting as a point of comparison, and the Future Self, by providing material with which it can be formed and embellished.



4.3. Longitudinal Developments in Self-Concepts Over Time

Research Question 3:

How do learners' motivational self-concepts develop over time?

One of the major shortcomings of previous motivation research to date is the lack of longitudinal developments in learners' motivational selves (Dörnyei, 2015; Al-Hoorie, 2018). To expand upon previous research, survey instruments in this study were administered at two timepoints, once at the beginning and once at the end of the school year, and participants' responses were compared. The surveys provided a direct measures of changes between timepoints. Since it was only possible to conduct interviews with students once at the end of the year, comparative qualitative data sources are not available; whilst learners did sometimes talk about changes in their self-concepts in the interview data, these self-reports were not conclusive. RQ 3 is thus answered via close inspection of the survey data.

4.3.1. Preliminary Analyses

4.3.1.1. Attrition Bias

In total, 270 students participated in this study. Of these 265 took part at T1, 173 (65.3%) of these continued to T2, along with 5 new participants who joined the T2 sample. To determine whether there were any notable areas of difference between the retained sample, and the 92 students who were no longer participating (the attrition group), descriptives for both sets of learners were compared. The retained sample and attrition group do not differ substantially in terms of learner characteristics, and it was determined that there were no systematic differences between the responses of those

who continued with and those who left the study. (See Section 4.2.1.1.2. for full overview of analysis of the retained and attrition group).

4.3.2. Comparison of Average-Score Means Over Time

4.3.2.1. Full Sample

Visual inspection of the average-score (averaged survey item means for each self-concept) means at T1 and T2 reveals an unexpected pattern. For eight of the nine factors, the mean scores at T2 are lower than at T1, as illustrated in Figure 4.17. The only exception is MSA CS for which the mean is the same at both timepoints ($M_{T1}=3.00$, $SD_{T1}=1.30$; $M_{T2}=3.00$, $SD_{T1}=1.42$). Parametric tests were run to establish whether any of these differences were significant. Since three t-tests were run simultaneously, a Bonferroni adjusted p-value of .017 ($=.05/3$) is employed as the cut-off level for significance. The only factor for which there was a significant change with this adjustment in place was Classical CS (outlined in bold in Figure 4.17), for which the mean at T1 ($M_{T1}=3.56$, $SD_{T1}=1.19$) was significantly higher than the mean at T2 ($M_{T2}=3.26$, $SD_{T2}=1.34$) with a difference of 0.329 ($SD=1.56$, 95% CI [0.09, 0.56], $t(170)=2.762$, $p=.006$, $d=.21$), though the effect size was small.

This means, contrary to expectations, that there were no significant changes in learners' self-concepts related to any variety of Arabic over the course of the academic year, except for their CS concept in Classical Arabic which actually weakened, indicating that they felt less capable in their skills of accessing religious texts. This is discussed further in Section 5.3.1 of the next chapter.

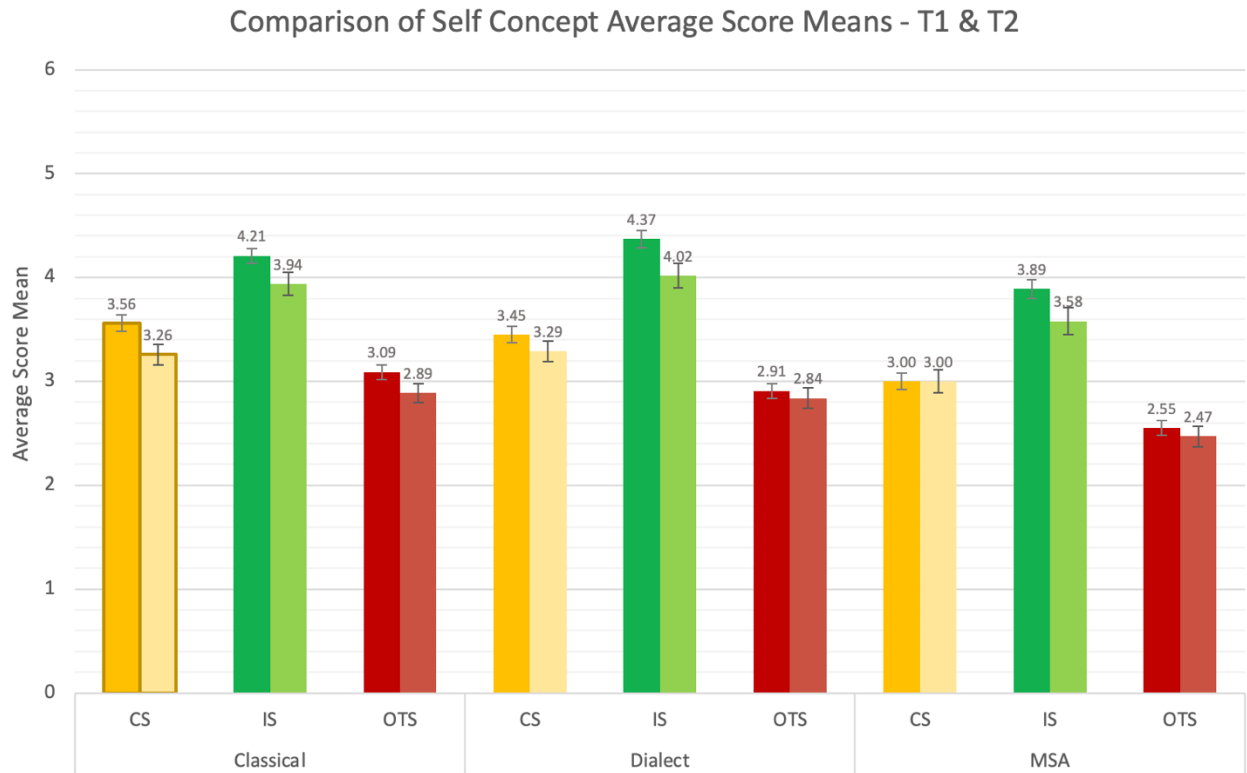


Figure 4.17. Bar graph for comparison between average-score means for self-concepts at T1 and T2.

(The first column for each variable shows the T1 score, followed by the T2 score.)²⁶

4.3.2.2. Gender

The next important question is whether these patterns emerged within subgroups (gender and heritage background) in the data when these are examined separately. When split by gender (M=52; F=112), employing a Bonferroni-adjusted p-value of .017 ($=.05/3$), the t-test results indicate that there are no significant differences for girls or boys between timepoints, in other words there was no evidence for any change between timepoints for either gender (See Appendix C.24 for full results).

4.3.2.3. Heritage Background

With the same post-hoc adjustment as above, the longitudinal differences were also explored when the data was split by heritage background (see Appendix C.25 for full t-test results). As can be observed from Figure 4.18, scores for each self-concept for NHLs

²⁶ 95% Confidence Intervals for each variable represented by error bars.

fell between the two timepoints. For HLs, each of them rose, albeit minimally. Though only five of these changes were significant, this highlights an interesting pattern: NHLs' self-concepts appear to deteriorate overtime, whereas HLs' appear to have improved.

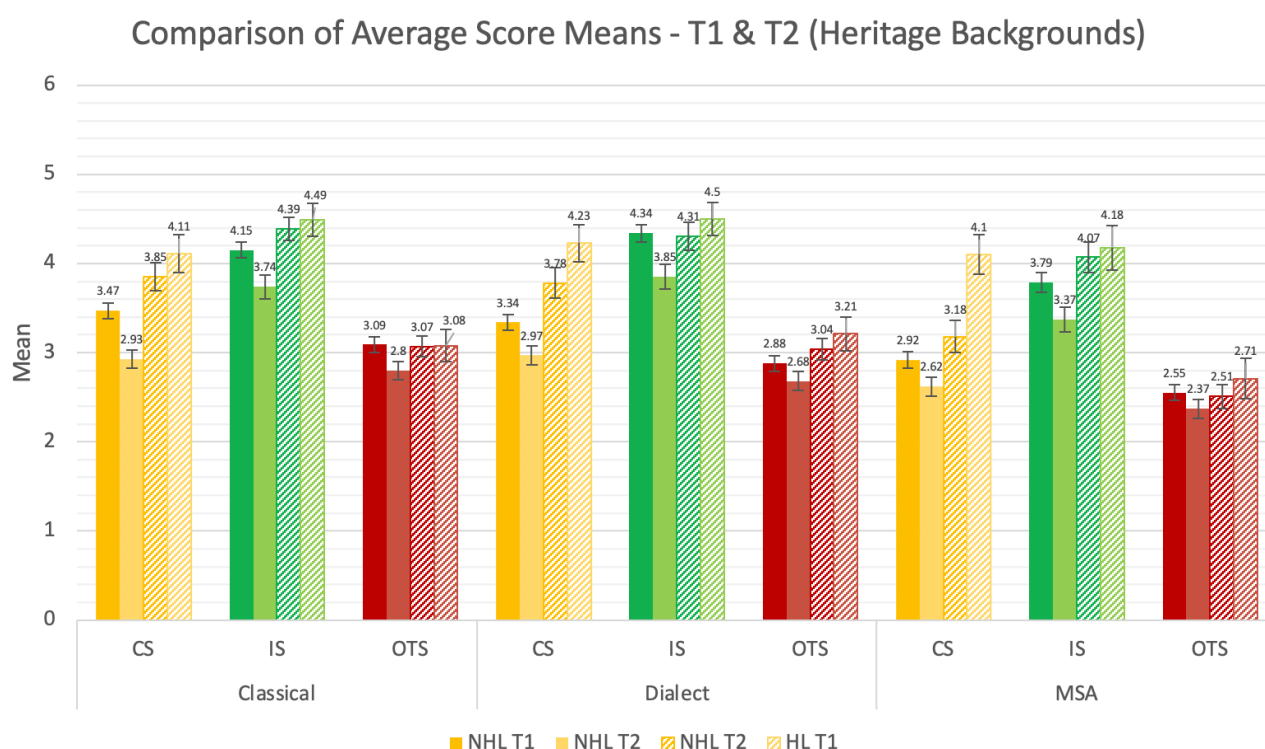


Figure 4.18. Bar graph for comparison between average-score means for self-concepts at T1 and T2 for HL and NHL subgroups. (The first column for each variable shows the NHL T1 score, followed by the NHL T2 score. This is followed by the HL T1 score, then the HL T2 score.)²⁷

Within the NHL group, there were four significant changes in their self-concepts over time. For Classical Arabic, the Current and IS concepts of learners in this group were significantly lower at T2 than at T1. Their CS concepts fell from over the midway mark on the Likert-scale ($M=3.47$, $SD=1.21$) at T1 to just under ($M=2.93$, $SD=1.16$) at T2 ($M=0.54$; $SD=1.51$, 95% CI [0.27, 0.81], $t(125)=4.013$, $p<.001$ $d=.36$). NHL learners' IS concepts also fell over time ($M=0.41$; $SD=1.83$, 95% CI [0.09, 0.73], $t(126)=2.526$, $p=.013$, $d=.23$). The majority of the schools in the sample are Muslim faith schools which incorporate Arabic into their curriculum as part of their Islamic ethos. Therefore, it can be assumed that the implicit, and often explicit aim of offering Arabic in the curriculum is to promote access to religious texts and to enhance students'

²⁷ 95% Confidence Intervals for each variable represented by error bars.

command of the language for use in relation to their faith. It is, as aforementioned, the Classical variety that is used in this domain. Paradoxically, rather than improve learners' perceptions of themselves in this domain, both current and future, the effect of the GCSE in this sample, it could be argued, has instead weakened them. If a key component of instruction in relation to motivation is the strengthening of, at the very least, learners' future self-concepts, then current GCSE Arabic provision appears to be failing NHLs in this regard.

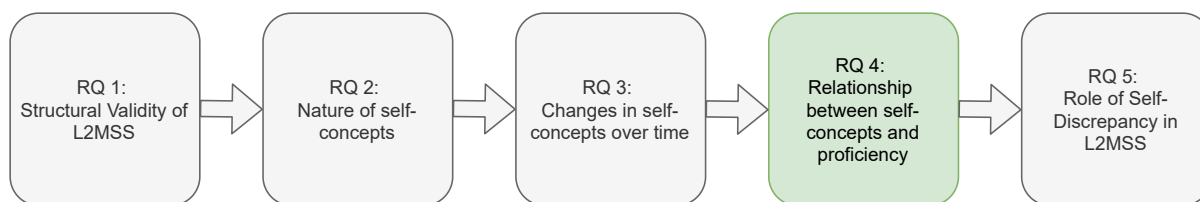
Given that the self-stated aims of the GCSE are to improve learners' abilities to communicate with native speakers, one would expect self-concepts related to this, namely Dialect (operationalised as the '*ability to communicate fluently with native speakers*' in the survey-scale opening statement) to improve. For HLs, there was no evidence of significant changes in self-concepts for this variety. This is perhaps because learners are aware that the language of the GCSE content is not the same as that used in their homes and therefore does not have a significant impact on their spoken communication skills. For NHLs, the evidence suggests that a year of studying GCSE Arabic had a negative impact on their Current and IS concepts in this language domain, with both decreasing significantly over time. For CS, the average-score mean fell from 3.34 (SD=1.27) at T1 to 2.97 (SD=1.22) ($\Delta=0.38$; $SD=0.60$, 95% CI [0.10, 1.66], $t(125)=2.655$, $p=.009$, $d=.24$), and for IS the drop was larger ($\Delta=0.49$; $SD=2.00$, 95% CI [0.14, 0.84], $t(125)=2.766$, $p=.007$, $d=.25$).

CS concepts for the MSA variety increased significantly for HLs between T1 and T2 ($\Delta=-0.92$; $SD=1.64$, 95% CI [-1.44, -0.40], $t(140)=-3.596$, $p=.001$, $d=-.35$). It seems, as a result, that a year's GCSE study only strengthened the current motivational self-concepts of those from heritage backgrounds, though minimally given the small effect size, and did not improve their future self-concepts.

4.3.3. Summary

It could be argued that exposure to a new curriculum, drastically different in pace and content from KS3, during which most of these students were studying Arabic from teacher-generated syllabi and materials, could have an initial weakening effect on

learners' self-concepts. This is alluded to in discussions on learners' past learning experiences with Arabic in the interviews, explaining the low to middling average-score means at T1. What is striking is that there were no positive significant changes in these scores over the course of the year for any self-concept except for HLs' MSA current selves. This is concerning on two fronts. The first is that the first year of the GCSE does very little to motivate learners, in the sense of promoting IS images and/or decreasing the gap between the Current and Ideal selves. In fact, for many NHLs it does the opposite in the realms of classical and colloquial skills. Secondly, in the one domain which could have a significant impact on eventual academic performance (MSA CS), HLs are advantaged over their NHL counterparts, reinforcing the view that the GCSE is not equally accessible to all learners regardless of background.



4.4. The Relationship between Self-Concepts and Proficiency

Research Question 4:

What is the relationship between proficiency and self-concepts

- a) Within timepoints*
- b) Between timepoints*

Both self-reported measures of proficiency and proficiency test data were gathered at T1 and T2. These data were examined at both timepoints to see whether or not there were any differences between sample sub-groups and whether the scores for the self-report scales and proficiency tests changed over time for the sample as a whole, as well as for the gender and heritage background subsamples. Interview data regarding specific skill proficiency was also identified and incorporated into this analysis in order to supplement and expand upon quantitative findings.

4.4.1. Full Sample Analyses

4.4.1.1. Proficiency Scores

The total possible score for the cloze test proficiency test was 25. For the Y/N vocabulary test, the highest possible score was 60.

4.4.1.1.1. Descriptive Analyses

There were no outliers for either the Y/N or cloze tests, or the total scores at T1 or T2. Whilst the skewness z-scores for the Y/N total T1 was slightly above the 3.29 cut off, the raw scores were well within the acceptable limits for parametric tests, which were subsequently employed.

4.4.1.1.2. *T-Tests*

At T1, the mean score for the cloze test for the entire sample was 13.74; in other words, 54.97% of the total possible score. For the Y/N test, the mean score was 24.2, a lower percentage of 40.29% of the total possible score of 60 (see Table 4.16). These are illustrated in Figure 4.19.

Table 4.16. Descriptive Statistics – Proficiency Scores T1 & T2

	N	Mean	% of Total	SE	SD	Skewness	Std. Error	z-score	Kurtosis	Std. Error	z-score
YN Total T1	262	24.1718	40.29	0.957	15.496	0.676	0.15	4.51	-0.464	0.3	-1.55
YN Total T2	174	25.31	42.18	1.147	15.131	0.416	0.184	2.26	-0.465	0.366	-1.27
Cloze Total T1	264	13.7424	54.97	0.384	6.245	0.14	0.15	0.93	-0.838	0.299	-2.80
Cloze Total T2	170	15.46	61.84	0.519	6.773	-0.416	0.186	-2.24	-0.701	0.37	-1.89

What is surprising, from a preliminary examination of these mean totals, is just how little they increased over the course of the academic year. When compared using a paired t-test, the scores at T2 did not emerge as significantly higher than those at the first timepoint (see Appendix C.26 for results). This is addressed further in the Discussion Chapter (Section 5.3.2).

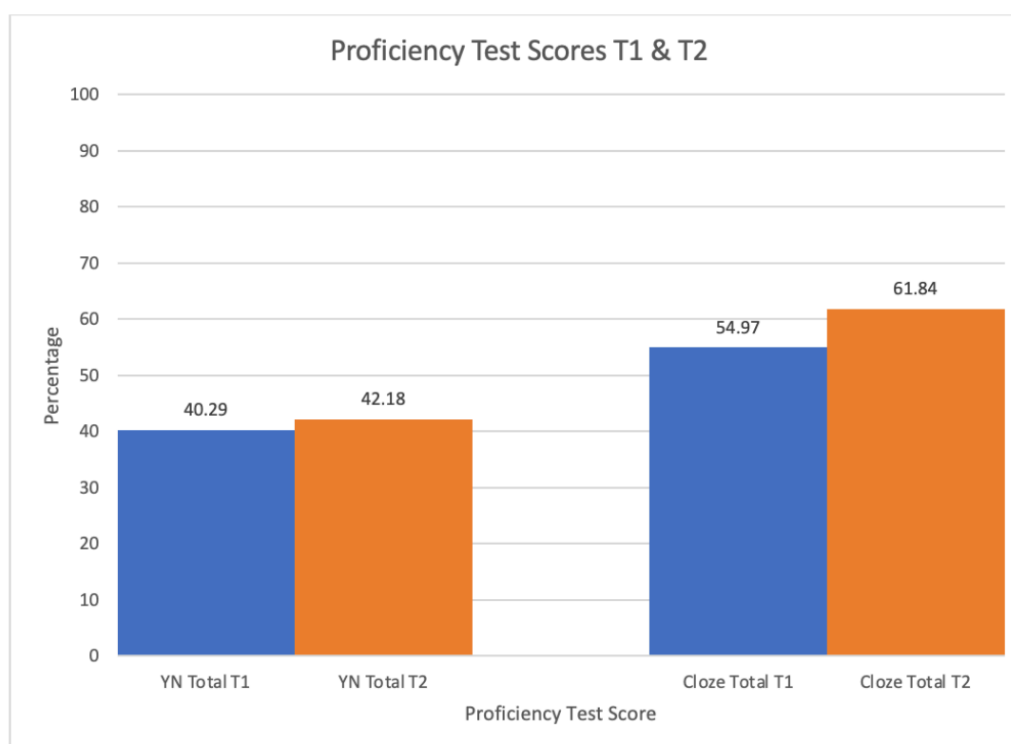


Figure 4.19. Bar graph for comparison of proficiency test scores at T1 and T2.

4.4.1.2. Self-Rated Proficiency

Self-rated proficiency scores covered the following eleven skills, which spanned the three different varieties in line with their contexts of use:

- MSA
 - Writing
 - Speaking
 - Reading Comprehension
 - Listening Comprehension
 - Phonologic decoding (general)
- Dialect
 - Speaking
 - Listening Comprehension
 - Ability to distinguish between dialect and standard Arabic
- Classical
 - Reading Comprehension
 - Listening Comprehension
 - Phonologic decoding (with diacritics)

4.4.1.2.1. *Descriptive Analyses*

Learners were asked to rate their proficiency within different skill domains associated with the three Arabic varieties. Mean scores for T1 (yellow) and T2 (blue) are illustrated in Figure 4.20. Visual inspection of the mean scores shows that learners rate their proficiency in MSA-related skills higher than both Dialect and Classical at T1, with the exception of Word Reading with *Harakat* (diacritics). This may be explained by the comparative ease of reading with vowel markings, plus the familiarity with the Quranic script which invariably includes these, which learners emphasised in their interviews). The same pattern is also apparent at T2. This appears to contradict the findings regarding learners' CS concepts which showed MSA to be the lowest of all three varieties. This discrepancy is possibly explained by the fact that the self-rated proficiency scores for MSA also included MSA Speaking and Listening (since these are

components of the GCSE exam), whilst the way in which MSA was operationalised in the scale-items opening statement was confined only to reading written sources. Nevertheless, the mean scores for MSA Reading comprehension were still, upon first inspection, higher than those of the classical and dialect skills. Another possible explanation is that inclusion of *confidence* scores in the CS concept, which it can be deduced must be lower for the MSA variety.

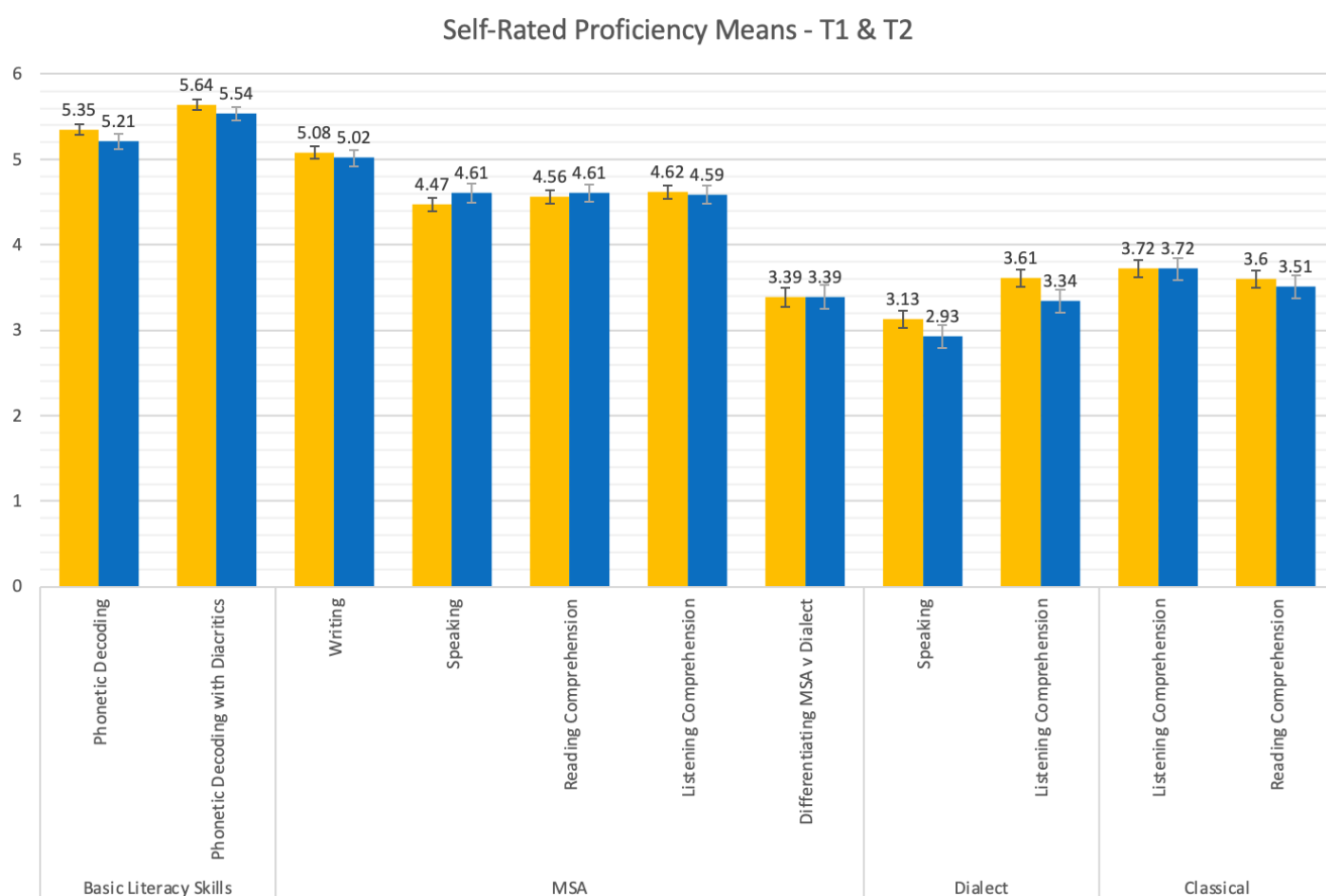


Figure 4.20. Bar graph for comparison of Self-Rated proficiency scores at T1 and T2.²⁸

4.4.1.2.2. Normality and Outliers

The skewness and kurtosis scores for numerous items fell outside the acceptable limits for both the raw scores and z-scores. The extreme scores for skewness and kurtosis were particularly noticeable for items involving word reading at both timepoints. This is unsurprising, given that the majority of students had studied Arabic for at least a year

²⁸ 95% Confidence Intervals for each variable represented by error bars.

before beginning the GCSE, during which time they would likely have developed word reading skills. These two items were therefore ignored from analysis, given their highly negative skew and large kurtosis scores. A number of the remaining items also exhibited skewness and kurtosis values which exceeded acceptable limits for parametric tests, and thus non-parametric tests were utilised in order to analyse the relationships between these scores.

Once the items concerning word reading were removed from the dataset, only one item contained outlier responses. This was MSA Writing, for which 4 students gave a response of 1, which resulted in a z-score of -3.46, just outside the acceptable 3.29 cut off limit. Since non-parametric tests are considered to be robust to these infringements for mean comparisons (Field, 2015), these scores were not removed or corrected.

4.4.1.2.3. *Mean Comparison - Changes over Time*

There were no significant changes in learners' self-rated proficiency between T1 and T2 when the sample was examined as a whole, i.e., learners did not regard themselves as having improved over the academic year in any of the domains to which the items referred. Test results are provided in Appendix C.27.

4.4.2. Gender Differences

4.4.2.1. Proficiency Tests

Scores for males and females appear to be very close for the cloze tests, with only a slight advantage for male students in the vocabulary test at T1, which explains their slightly higher combined mean scores. The standard deviations are also very similar for the two groups, suggesting similar ranges of scores amongst both male and female participants (see Appendix C.28).

4.4.2.1.1. *Mean Comparisons*

Despite the fact the previous research indicated girls are more motivated to learn languages than boys (Carreira, 2011; Henry, 2010), no significant differences in actual

proficiency were present between genders in this study at either timepoint. There were also no significant changes between the timepoints within either gender, providing no evidence in this data, therefore, that gender has an effect on proficiency development (see Appendix C.29 and C.30).

4.4.2.2. Self-Rated Proficiency

Mean scores for the SR Proficiency variables separated by gender are presented in Figure 4.21. Males exhibited a slightly higher mean than females for all SR proficiency items at both timepoints, with the exception of MSA Speaking T1 and Quran Listening Comprehension at T2. The most noticeable difference between the two groups, based on a visual inspection of the descriptive data, is reported at T2 for MSA Speaking (see Appendix C.31 for full t-test results).

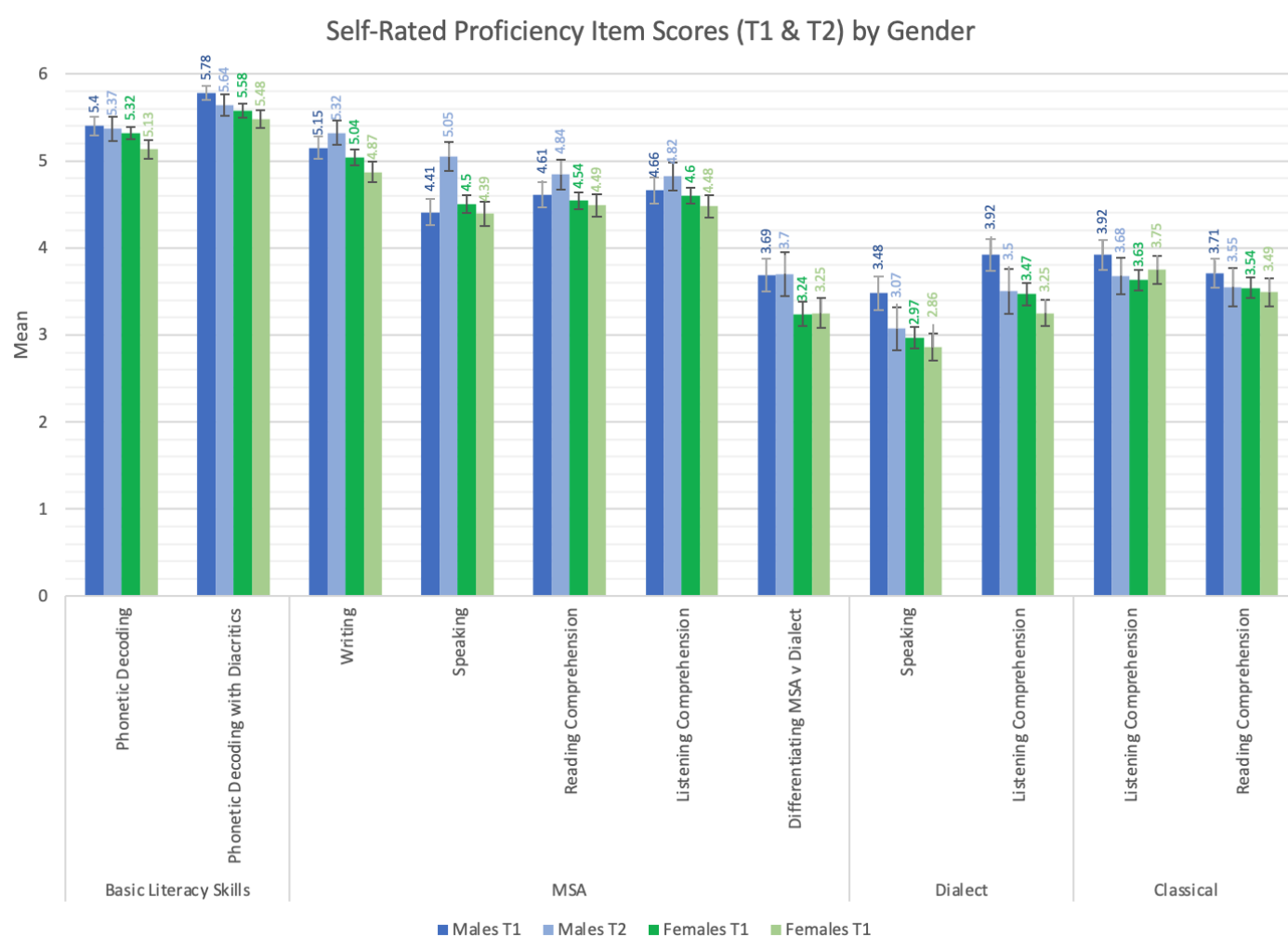


Figure 4.21. Bar graph for comparison of Self-Rated Proficiency scores at T1 and T2 separated by gender.²⁹

4.4.2.2.1. Mean Comparisons

Since nine tests were run simultaneously, an adjusted significance level of .005 was implemented. The results of independent samples Mann-Whitney U Tests indicated that there were no significant differences between genders for any of the self-rated proficiency scales in this dataset (see Appendix C.32).

To examine whether or not the self-rated proficiency scores improved over time within on gender, a related-samples Wilcoxon Signed Rank test was run on the data split by

²⁹ 95% Confidence Intervals for each variable represented by error bars.

gender. The only significant difference between timepoints concerned male learners' self-rated proficiency scores for MSA Speaking which improved from $M_{T1}=4.41$, $SD=1.39$ to $M_{T2}=5.05$, $SD=1.30$ ($p=.020$), see Appendix C.33.

4.4.3. Heritage Background

4.4.3.1. Proficiency Tests

Visual inspection of the mean scores indicates that on both tests HLs scored higher at both timepoints than the NHLs (see Appendix C.34). The standard deviation statistics are also higher for HLs, suggesting a wider variance in scores amongst this group at both timepoints.

4.4.3.1.1. *Mean Comparisons*

The gap between HLs and NHLs' raw scores decreased from T1 to T2. However, at both timepoints, the scores for HLs are significantly higher than NHLs, both for the Y/N and the cloze tests, as confirmed by independent samples t-tests (see Table 4.17 and Figure 4.22; full results provided in Appendix C.35). If the Y/N test is to be taken as an indication of vocabulary knowledge, and the cloze test of broader proficiency, including grammatical knowledge, this offers evidence of an advantage for HLs in both these skill domains. The effect sizes were medium to large.

Table 4.17. Independent T-Test results for comparison between HLs and NHL proficiency test scores at T1 & T2									
		N	Mean	SD	SE	t	df	Sig.	Effect size (for HL diffs)
Y/N Total T1	NHL	197	19.23	11.70	0.83	-9.263	87.1	.000	1.43
	HL	65	39.14	16.06	1.99				
Y/N Total T2	NHL	126	22.10	12.82	1.14	-4.119	67.8	.000	.77
	HL	48	33.75	17.47	2.52				
Cloze Total T1	NHL	199	12.72	5.52	0.39	-4.213	89.4	.000	.65
	HL	65	16.86	7.26	0.90				
Cloze Total T2	NHL	124	14.48	6.43	0.58	-3.366	168	.001	.54
	HL	46	18.13	7.03	1.04				

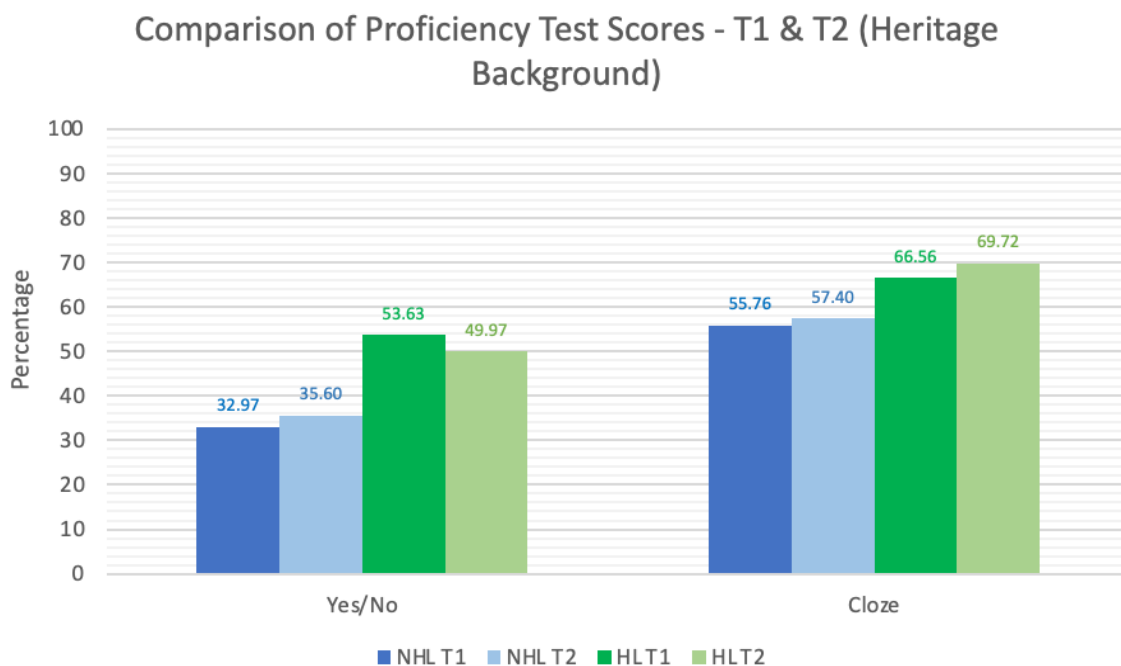


Figure 4.22. Bar graph for comparison of proficiency test scores at T1 and T2 separated by heritage background.

Despite the decrease in the gap between the two groups by T2, no significant longitudinal improvements were apparent in either HL or NHL scores once Bonferroni adjustments to the p-value were made (.05/2 tests=.025). Though HLs began with and maintained higher proficiency scores at T1 and T2, there is no evidence to indicate that heritage background has any effect on longitudinal improvements (see Appendix C.36). This finding is discussed in Section 5.3.2.

4.4.3.2. Self-Rated Proficiency

Visual inspection of the descriptive statistics indicates that means for self-rated proficiency scores are higher across the board for HLs (see Appendix C.37 for full descriptive statistics). The biggest area of difference, from a cursory inspection, lies in learners' ability to differentiate between MSA and Dialect at T2 (see Figure 4.23). This is followed by their Dialect speaking and comprehension skills. This is not unexpected given that HLs are likely to have more exposure to and knowledge of the dialect, and also given that an extra year of this exposure combined with school learning seems to have increased this skill set amongst these learners.

4.4.3.2.1. *Mean Comparisons*

With the adjusted significance level of .005 imposed, the results of independent samples Mann-Whitney U Tests demonstrate that there are also no significant differences between HLs and NHLs for any of the self-rated proficiency scales at T1 (Appendix C.38). This indicates that, although HLs outperform NHLs in the proficiency tests, there is no evidence in this data that HLs' own perception of their skills differs from their NHL counterparts, even when it comes to skills involving dialect. However, by T2, HLs score higher on all SR proficiency variables.



Figure 4.23. Bar graph for comparison of Self-Rated proficiency scores at T1 and T2 separated by heritage background.³⁰

Learners scores at T1 and T2 were also examined by group for any significant longitudinal changes. With the Bonferroni-adjusted p-value in use, no significant longitudinal changes were noted in either HLs' or NHLs' self-rated proficiencies (see Appendix C.39).

4.4.4. Path Model Analyses

The majority of studies investigating the L2MSS face limitations in terms of: (a) their use of only self-reported outcome measures concerning intended effort, and (b) their lack of examination of the longitudinal relationship between self-concepts and outcome measures. To address these shortcomings, the relationship between self-concepts and

³⁰ 95% Confidence Intervals for each variable represented by error bars.

multiple proficiency measures both within and between timepoints is examined in this study.

In order to assess the relationships between different self-concepts and proficiency across timepoints, a series of path models were run, one for the combined 'Overall' higher-order construct scores for all varieties and one per language variety. The advantage of path modelling, in comparison to simple regression, is that it allows for the investigation of complex relationships between variables (e.g., Blunch, 2013; Kline, 2015; Tabachnik & Fidell, 2014). The original analytical approach had been to run a Structural Equation Model (SEM) for this data. SEM analysis incorporates measurement modelling with path modelling, which uses variances and covariances to explore linear relationships between dependent and independent variables. However, combining a CFA with path modelling within this model meant the incorporation of a large number of parameters, which the current sample size could not support. One widely used rule of thumb is that 5-10 observations are required for each parameter estimate (e.g., Kline, 2015), although there is no consensus on the sample size required for SEM. 36 parameters were estimated between the latent constructs, without the incorporation of the item regressions onto the latent variables. Since this alone would require 180 observations according to the above rule of thumb (even at its lower bound of 5 observations per parameter), the dataset was not considered sufficiently large to run the traditional SEM models.

A common solution to this issue is to split the SEM analysis into its component parts and use the factor scores generated from the CFA to run the subsequent path models (Devlinger & Rosseel, 2017). This approach is henceforth referred to as Factor Score Path Modelling, or path models.

Since the path modelling function in R requires a complete dataset, missing data from the motivation scale items were imputed (see Section 4.1.1.1.1). This data was used to run a confirmatory factor analysis from which the factor loadings for each item were derived. A CFA was run for each Arabic variety to identify whether the items loaded onto the three hypothesised constructs of CS, IS, and OTS. Following this, a higher-order CFA was conducted to examine whether the latent self-concepts from each variety also

loaded onto larger, overall Arabic motivational self-concepts. All items loaded onto their underlying hypothesised latent factor, with the exception of the *IS Plausibility* item which cross loaded onto both the IS and CS concepts.

Upon generation of the factor loadings for the specified model for each variety and for the Overall model at both timepoints, the values for each latent variable were transformed into their standardised factor score equivalents. These were used to run the subsequent path models.

4.4.4.1. Self-Rated Proficiency Path Models

Path Models were first run by inputting the data for the self-concept variables along with the factor scores for self-rated proficiency to examine the relationship between these constructs per language variety, and the Overall model.

This was run as a saturated model, in which the number of estimated parameters is equal to or greater than the number of data points; in other words, there are zero degrees of freedom. In these cases, goodness-of-fit is determined by comparing one or more of the following parsimony-adjusted measures of fit: Loglikelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC) indices, which are presented in Table 4.18. Since there are no guidelines on acceptable ranges for these indices, models are assessed comparatively; the lower these scores, the better the model fits the data (Field, Field & Miles, 2012). In this case, the best model fit indices emerge in relation to the Overall model, followed by the MSA model.³¹

³¹ Since root mean square error of approximation (RMSEA) and other fit indices are unavailable to determine exactly how well the model fits the data, these models should be interpreted with some caution.

Table 4.18. Model Fit Indices for Language Variety Models and Overall Model – Self-Rated Proficiency									
Variety	Free Parameters	Observations Used	Test Statistic (Robust)	Degrees of Freedom	Scaling Correction Factor	Loglikelihood	AIC	BIC	Sample Adjusted BIC
Overall	36	149	657.342	28	1.102	-1092.62	2257.24	2365.38	2251.45
MSA	36	149	502.778	28	1.174	-1439.67	2951.35	3059.49	2945.56
Dialect	36	149	466.691	28	1.171	-1535.99	3143.97	3252.12	3138.19
Classical	36	149	511.193	28	1.193	-1459.55	2991.11	3099.25	2985.32

Emergent path models are illustrated in Figure 4.24. What is noteworthy, and to be expected if SR Proficiency is considered a complementary measure of the CS, is that these two variables interact reciprocally overtime when the scores for all language varieties are combined. When the scores for varieties are separated it becomes clear where the sources of this interaction stem from. SR Proficiency at T1 predicts CS for MSA and Classical at T2, whereas CS T1 predicts SR Proficiency at T2 for Dialect only. Contrary to hypothesised models which hold that the IS wields predictive power, none of the varieties, nor the combined model, exhibit any link between the IS at T1 and SR Proficiency as the outcome measure at T2. For the Classical variety, SR Proficiency appears to have significant but weak predictive power in regard to the IS at T2, but this relationship disappears when the varieties are combined.

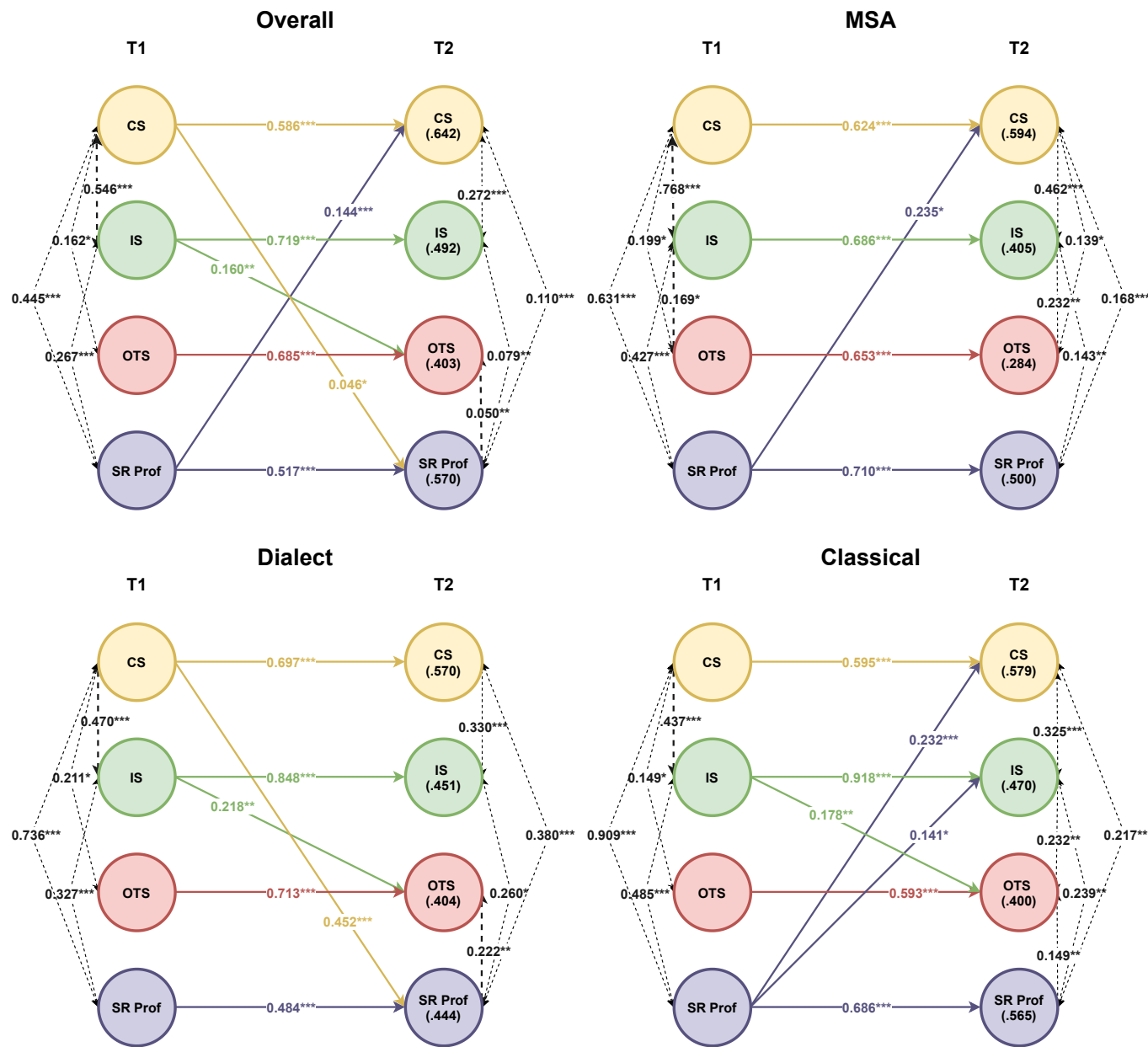


Figure 4.24. Factor Score Path Models for each language variety and Overall model
Key: CS = Current Self; IS = Ideal Self; OTS = OTS; SR Prof = Self-Rated Proficiency

4.4.4.2. Proficiency Test Path Models

A second set of path models were run with the proficiency test scores. The self-concept latent variables were entered along with proficiency scores for the Y/N and cloze tests individually. This decision was made on the basis of the assumption that these two scores appear to tap into different skill sets. Thus, rather than combine them into one proficiency variable, they were retained as distinct scores to examine their independent interactions with the self-concept variables.

In line with the L2MSS, the hypothesised model predicts that self-concepts at T1 contribute to the development of both self-concepts and proficiency at T2. There is also predicted to be a relationship between proficiency at T1 and self-concepts at T2, though this is not directly addressed by descriptions of the L2MSS to date.

4.4.4.2.1. *Overall Arabic Selves*

The Overall higher-order constructs model for Arabic, illustrated in Figure 4.25, reveals some interesting findings. Contrary to the hypothesised model, self-concepts at T1 have little predictive strength to any variables other than their own counterparts at T2. Only the IS factor contributes in relation to a self-concept other than itself. Given studies indicating the promotive function of IS when it comes to outcome variables such as intended effort (e.g., Taguchi et al., 2009; Dörnyei, 2015; Teimouri, 2017), one would expect that a similar relationship would exist between IS T1 and CS T2. Instead, this study finds that IS significantly predicts variance in OTS at T2. In other words, the higher a person's IS at T1, the higher their OTS at T2.

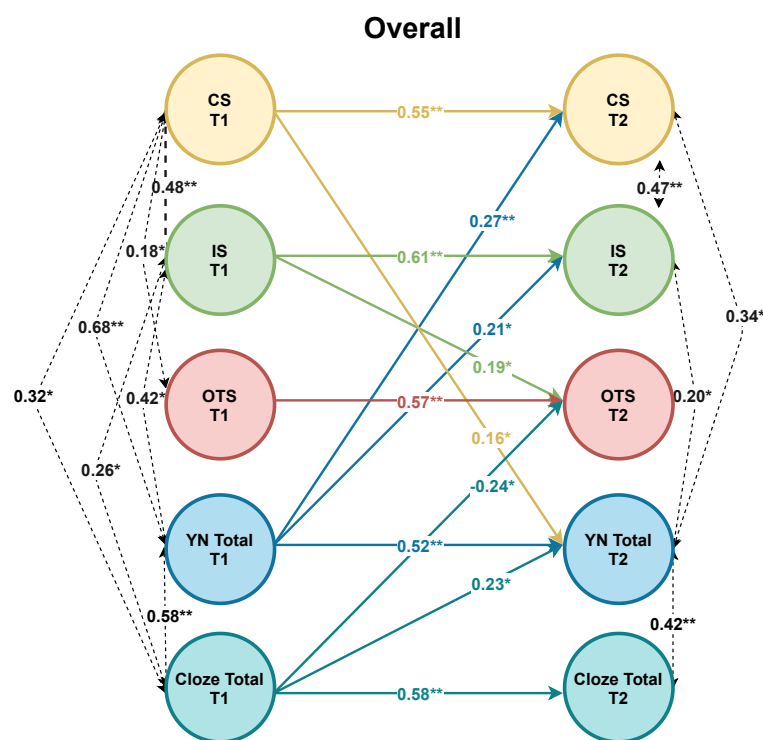


Figure 4.25. Factor Score Path Model with Proficiency Test variables – Overall Model

There was also a weak but significant predictive relationship between CS at T1 and proficiency at T2, though only the pathway between the CS and the Y/N variable emerged as significant. This lends weight to the idea that these two proficiency tests, while highly correlated, draw upon different skills which interact independently with different motivational selves. According to this path model, the higher a person's CS at T1, the higher their vocabulary knowledge at T2. At the same time, this CS has no measurable impact on the grammatical knowledge. The variables in this model explain noticeably more of the variance for the CS than the other factors at T2 (see Table 4.19).

Table 4.19. R2 Estimates for Model 1	
Variable	R ² Estimate
Current Self T2	0.639
Ideal Self T2	0.493
Ought-To Self T2	0.417
Y/N Total T2	0.435
Cloze Total T2	0.493

Though self-concepts at T1 appear to have little impact on proficiency at T2, the reverse stands out as being the case. There are significant paths between proficiency, of one type or the other, to each of the self-concepts at T2. Both CS and IS are positively

predicted by Y/N scores, whereas a negative relationship is evident between the cloze test and the OTS scores. It can thus be extrapolated, tentatively, that the higher one's vocabulary knowledge at T1, the higher one's CS and IS at T2. It is logical to assume that a higher level of proficiency at T1 would contribute to stronger CS concepts at both timepoints. What is more interesting however, is that a) this objective measure of proficiency at T1 predicts IS at T2, whereas learners' CS concept at T1 does not, and b) that the IS at T2 is only predicted by the T1 vocabulary knowledge test scores and not the T1 cloze test scores which draw upon learners' knowledge of grammar as well as vocabulary.

To further add to this complex picture of the interplay between proficiency types and self-concepts, cloze test scores at T1 emerge as a significant predictor of the OTS at T2, this time with a negative coefficient. A possible explanation for this is that less proficient students may feel greater pressure to avoid negative consequences.

4.4.4.2.2. *Individual Language Varieties*

Since analyses for individual varieties support the theory of their discreteness, the data for each variety was entered into the model to identify whether there were any differences at this level. As with the SR Proficiency model, the proficiency path models were saturated, thus the Loglikelihood, AIC, and BIC scores are compared to determine goodness-of-fit. The Fit Indices indicate that the model was best suited to the combined data for Overall Arabic, followed by the data for the Classical self-concepts, with the poorest fit emerging for MSA according to all three measures (see Table 4.20).

Table 4.20. Model Fit Indices for Language Variety Models and Overall Model – Proficiency Tests

Variety	Free Parameters	Observations Used	Test Statistic (Robust)	Degrees of Freedom	Scaling Correction Factor	Loglikelihood	AIC	BIC	Sample Adjusted BIC
Overall	55	147	662.24	45	1.214	-1604.14	3318.28	3482.76	3308.705
MSA	55	147	645.17	45	1.213	-1738.29	3586.58	3751.06	3577.005
Dialect	55	147	606.14	45	1.218	-1708.68	3527.37	3691.84	3517.790
Classical	55	147	612.04	45	1.231	-1633.05	3376.10	3540.58	3366.527

MSA

In many ways, the MSA model (illustrated in Figure 4.26) mirrors the Overall higher-order constructs model. The CS still emerges as a predictor of proficiency (YN) at T2. Similarly, the pattern whereby changes in self-concepts at T2 are driven by proficiency at T1, is maintained for this variety.

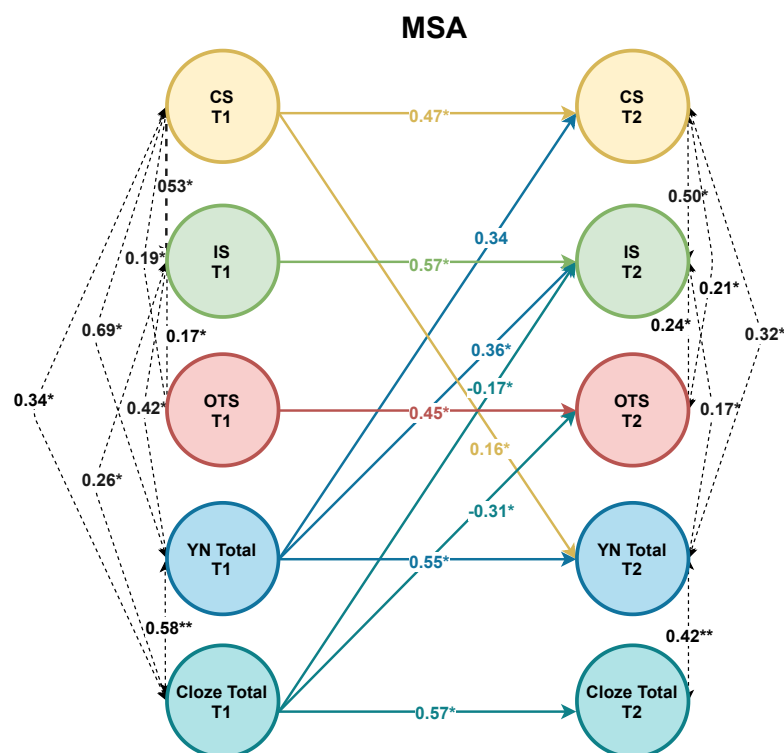


Figure 4.26. Factor Score Path Model with Proficiency Test variables – MSA Model

However, there are also notable areas of divergence. It appears that the IS for MSA at T2 is also predicted by the cloze test scores at T1, in addition to the Y/N scores. Interestingly, as it is for OTS, the regression coefficient is negative (-.17). This means that the higher a learner's score on the cloze test at T1, the lower their IS score at T2. Whereas it is possible to explain this relationship when it comes to OTS as more proficient students feeling less pressure to avoid negative consequences, when it comes to IS, the explanation is less transparent. Given that this relationship emerges only for MSA, one hypothesis is that learners with higher language abilities at T1 realise the limited application of MSA in everyday life, and are thus less likely to visualise themselves using it in the future.

At the same time, the R-square statistic for the IS is slightly lower than for Overall Arabic (see Table 4.21), with 44% of the variance explained, in comparison to 49%.

Table 4.21. R-Square scores for All Models				
Outcome Variable	Overall	MSA	Dialect	Classical
Current Self	0.639	0.616	0.568	0.574
Ideal Self	0.493	0.438	0.457	0.484
Ought-to Self	0.417	0.324	0.415	0.415
Y/N Total	0.435	0.429	0.436	0.436
Cloze Total	0.493	0.491	0.493	0.496

In addition, the OTS at T2 is no longer predicted by IS at T1, and the R-square score indicates that 32% of its variance is explained, in contrast to the 42% explained in the Overall model and the other varieties. This is the lowest explained variance of any of the constructs included in the models. Why is there no predictive relationship between IS and OTS over time for MSA exclusively? Why does it depend far more on the nature of the other self-concepts at T2 and the cloze test scores to a greater degree? It appears that there are other factors influencing the variance of the OTS factor besides those included in the model. These will need to be investigated further in future studies.

Upon examination of the correlations between variables at T2, it appears that IS and OTS are significantly correlated with a coefficient of 0.24, a relationship which does not emerge in the Overall Arabic model. CS is also significantly correlated with OTS at T2

(0.21). This suggests that self-concepts for MSA at T1 are closely related, and that this relationship contributes to the variance explained.

Dialect

For the Dialect variety, a sparser model emerges. In contrast to the four paths between proficiency at T1 and self-concepts at T2, for Dialect only two significant paths emerge (see Figure 4.27). In this model, CS is predicted by Y/N scores, and OTS is again negatively predicted by the cloze test scores.

Intriguingly, the IS is not predicted by any of the variables at T1, other than its own counterpart, though it is correlated, less strongly, with CS at T2 and both proficiency types. In comparison to MSA, the variance explained for CS is slightly higher, at 46%.

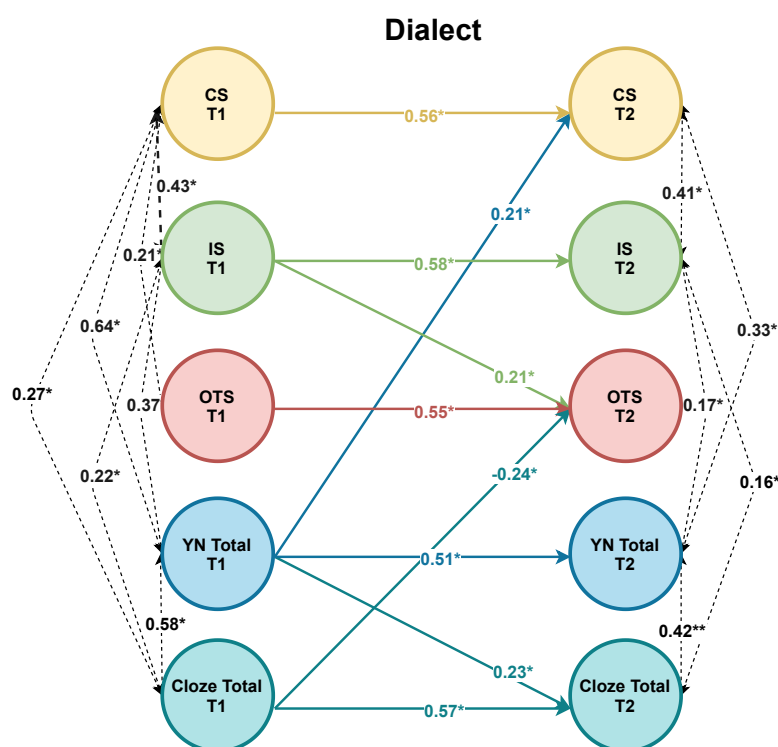


Figure 4.27. Factor Score Path Model with Proficiency Test variables – Dialect Model

That there is no significant path between the Y/N test (which tests vocabulary knowledge) and IS is unexpected. It indicates that vocabulary knowledge at the beginning of the year has no significant impact on learners' future self-images relating to speaking the language. One possible explanation for this is that learners are, or have

become, aware of the limited utility of the words they are learning in an everyday speaking context.

In contrast to both the Overall and MSA models, none of the self-concepts at T1 predict proficiency at T2. This is expected in the context of Dialect since the language used for spoken Arabic diverges from the content of the GCSE (and thus the proficiency tests which examine MSA). Whereas CS predicted Y/N proficiency in each of the other models, a learner's CS concept in the realm of Dialect appears to have no bearing on the development in their proficiency. This is counter-intuitive particularly in the context of HLs, since these learners should, in theory, have some advantage in proficiency due to their prior exposure to and knowledge of the language.

Classical

The Classical model mirrors most closely the Overall one, with the exception of a pathway between the T1 cloze test and the T2 Y/N test (see Figure 4.28). The CS variable at T2 has the highest R-squared score with 57% of the variance being explained, predicted predominantly by its counterpart at T1 (0.48), indicating the stability of the construct and by the Y/N Total (0.26).

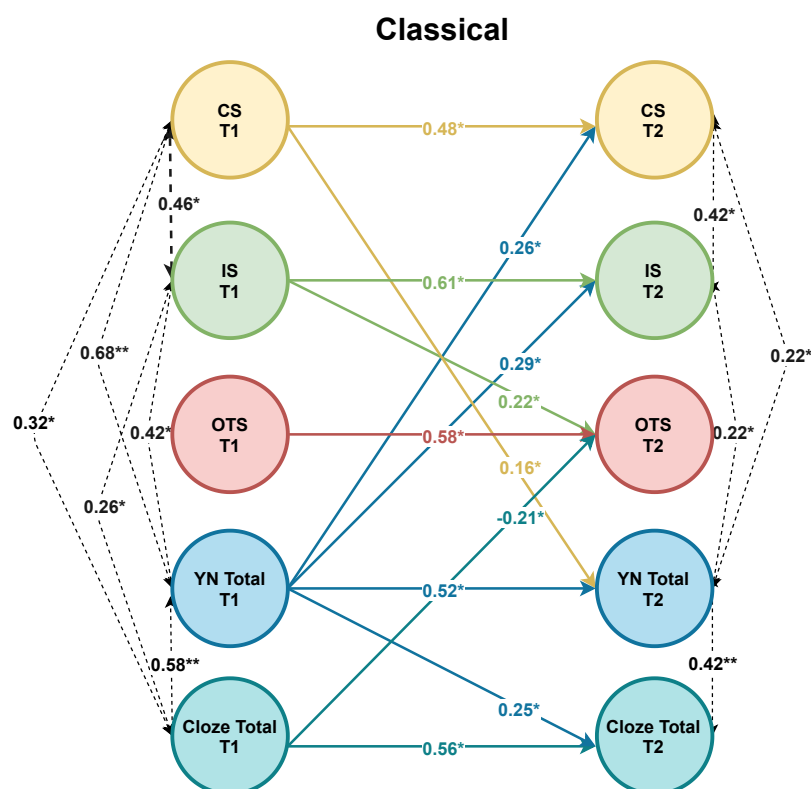


Figure 4.28. Factor Score Path Model with Proficiency Test variables – Classical Model

The pathway between IS and OTS also emerges as significant, as is the case with Dialect and the Overall model.

4.4.4.3. Summary

Based on the findings of prior research (see Section 2.1.2.1.1) and the underlying model of motivation described in the L2MSS, the hypothesised relationship between self-concepts and proficiency outcome measures would entail a promotive function of self-concepts, particularly the IS, in regard to outcome measures over time. This model was not supported by the data in the present study. Rather, what could be termed a ‘proficiency-driven model’ emerged whereby each self-concept at T2 is predicated by different proficiency measures. Supporting the claim that learners interact differently with individual language varieties, the number and strength of the regression coefficients emerging for each model were different.

4.4.4.4. Following the Thread: Differences between Vocabulary and Grammar

The difference in regression paths emerging for each of the two proficiency tests is noteworthy. Whereas scores for the Yes/No vocabulary-centred test regressed onto the CS and IS in the Overall higher-order constructs model with positive coefficients, as well as the factor scores for each variety, scores for the cloze tests regressed, negatively, only onto the OTS. Simply put, this suggests that the higher one's vocabulary knowledge the higher their CS and IS; on the other hand, the higher one's grammatical knowledge, the lower one's OTS.

In order to explore possible explanations for this pattern, the interview data was revisited. References to *grammar*, *vocabulary*, and each of the four skills (*reading; writing; speaking; listening*) were coded using these *a priori* terms. The data points for the grammar and vocabulary codes were then sub-coded into the emergent categories in Table 4.22.

Table 4.22. Frequency of Grammar and Vocabulary Codes		
Main Code	Subcode	Frequency
Grammar	-	55
	Negative references	25
	Positive references	13
	Mosque instruction	7
Vocabulary	-	151
	Learning—negative	45
	Learning—positive	22
	Utility	18
	Lack of utility	6

4.4.4.4.1. Vocabulary

The issue of the volume of vocabulary constituted the majority of negative references to vocabulary learning, previously discussed in Section 4.2.6.1.2.3. What potentially alludes to the reason for the significant regression coefficients for CS and IS are a number of references to the utility of vocabulary learning. Multiple participants

expressed their ability to apply their vocabulary knowledge and the vision of being able to apply this knowledge in the future to real life situations, predominantly when reading Quran or when travelling.

Say, if I've learnt some words in Arabic, and then I'm reading the Quran, I'll understand, like, the topic.'

I think, like, the key words that I've learnt, I'll be able to use, like, if I go on holiday or something. I do know, like, the key words. So I'll be able to use them if I'm on holiday. And stuff like that.

To explore this issue further, a matrix was generated from the interview data, cross-tabulating the codes for grammar and vocabulary, with the four skills, the results of which are presented in Table 4.23.

Table 4.23. Cross-references to Grammar/ Vocabulary and Four Skills				
	Listening	Reading	Writing	Speaking
Grammar	0	1	4	2
Vocabulary	4	16	9	14

What is immediately noticeable is the frequency of references made to vocabulary in the context of Reading and Speaking, suggesting an awareness of a relationship between vocabulary knowledge and these two realms of Current and Ideal selves. The discussion above concerning the predominance of positive current and future self-concepts in regard to classical and colloquial Arabic appears to explain why higher Y/N test scores would predict stronger Current and Ideal Selves. These two domains overlap with the skills of reading and speaking respectively. Thus, it seems learners are aware of how they can apply their present and future knowledge in these realms, which are relevant to them.

4.4.4.4.2. Grammar

There were three times more references to vocabulary than to grammatical knowledge. Seven of the 55 mentions of grammar occurred in the context of mosque learning, which was consistently described as superior to school teaching methods in this respect. Whilst the number of mosque students in the sample is insufficient to run statistical

analyses into whether their OTS scores were statistically lower than the remainder of the population, this is an area for future research. Is it the case that those students who perform better in grammatical tests due to extra-curricular learning feel less obliged to learn Arabic since they are engaged in voluntary learning outside of school?

What is worthy of note is that 7 of the 13 positive references to grammar learning come from a single school. There are a number of unique features of this school in the context of language, the most pertinent of which is the fact that learners had learned, or were in the process of learning, other foreign languages, including Latin. The expressions of appreciation for grammar learning, it could be suggested, may either be a result of a transference of this attitude amongst students themselves from other language learning experiences, or the approach to teaching grammar by the Arabic teacher which makes this component appealing. The predominance of these references may also partially explain the apparent inverse relationship between grammatical knowledge and the OTS, given the fact that students in this school chose to study the subject and were not subject to pressures from school. This suggests the potential necessity of nesting the data in schools when running the Path Model analysis in order to control for between-school differences.

4.4.4.5. Following the Thread Further: Classroom-Level Effects

In the interview data a number of references to school-level variables were made, including the effects of teacher and teaching approach on learners' experiences with Arabic. Students in one class in particular mentioned the fact that their teacher had given them an entire lesson on the future applications of Arabic:

... He gave every class a whole lesson on what you can do with Arabic and then, that's helped us a lot. Because we realised that you can do a lot with Arabic. Because we thought that Arabic would be useless, once we get out of secondary education we'd bin it, but then he's changed some people

Learners in another school commented on the fact that they had experienced a high turnover of teachers and this had resulted in a lack of progress in the language:

Arabic is a new language and our teachers, they'd keep on leaving and stuff so we don't actually learn anything... We're just kind of going back to the basics every year and then restarting...

The existence of this theme within the data highlighted the necessity of examining school-level differences amongst other variables in both the qualitative and quantitative data.

4.4.4.5.1. Interviews

The first thread to be followed was the potential for the manifestation of school-level differences in learners' FS concepts. References to different categories of FS image were cross-tabulated with the 8 different schools participating in the interviews. Three schools emerged as having notable results in this domain, as illustrated in the findings in Appendix C.40. In the first, Class 5, the two predominant FS categories were *Religion* and *Travel*, both of which were identified by over half the students. Interestingly, this is the school from which the first above-mentioned quote was drawn, in which the teacher for the boys' class dedicated a whole lesson to possible future uses of Arabic. *Travel* was also identified by 5 out of 6 of the students in Class 16. This may well be explained by the fact that learners completed their GCSE just before the interviews had taken place, and therefore, as a result of this achievement, were eager to put this to use in the future. A final noteworthy figure is that 92% of learners in Class 8 remarked that they saw Arabic as playing no role at all in their futures. Not surprisingly perhaps, this is the same school from which the second quote mentioned above was taken, in which participants felt that their learning trajectories had been significantly disturbed by repeated changes in teaching staff.

The category of Feelings was also examined in order to provide some insight into learners' CS experiences. The majority of learners in Class 16 also expressed feelings of enjoyment in reference to their Arabic learning, though this is perhaps retrospectively explained through the relief of having completed the GCSE. The majority of learners in Class 13 also expressed overall positive feelings towards learning Arabic, explained potentially by the fact that they chose to study the subject, often alongside other languages, and had a teacher undergoing training and supervision by the MFL

department. A third of learners in Class 5 also expressed positive feelings towards Arabic, and in particular their teacher, who, as demonstrated by the above quote, was described as expending a large amount of energy and effort towards their teaching.

In the realm of the OTS, it is interesting to observe that the predominant source of obligation amongst Class 16 students stemmed from the prospect of their own sense of disappointment. This is expected given the pressure they had just undergone in order to complete the GCSE a year early in order to avoid the potential of having to sit the exam under the new specifications. In Class 8, where almost all students expressed no desire to integrate Arabic into their FS images, the most prominent OTS category was, unsurprisingly, School Pressure. Students in Class 17, on the other hand, appear to be most subject to Parental Pressure, with 67% of students identifying this in their interviews. Why students in this school in particular appear to be under more noticeable parental pressure in comparison to others is unclear. What can be drawn from the above, however, is that there are clear patterns of responses amongst students in different classrooms.³²

4.4.4.6. Intra-Class Correlations

The interview data suggests that classroom-level factors can have an impact on learners' motivational self-concepts and their proficiency scores. This is to be expected to some degree, given the integration of the (often overlooked) L2 Learning Environment into the L2MSS model. In order to examine the extent of the classroom-level variable on these constructs, Intra-class Correlation Coefficients (ICCs) were calculated in R using the ICC package (Wolak, Fairbairn & Paulsen, 2012). The ICC can be interpreted as 'the proportion of the variance explained by the grouping structure in the population' (Gelman & Hill, 2007, p.258). This index goes from 0, if the grouping conveys no information, to 1, if all observations in a group are identical. In other words, the ICC scores identify the proportion of variance between classrooms as a proportion of all the variance per construct. They can be interpreted as a percentage; for example,

³² Another interesting area of investigation was the consistency of references to different themes in the surveys and interviews across the sample. Details of this analysis can be found in Appendix C.41.

for MSA OTS at T1, the ICC is .10, meaning that 10% of the total variance is explained by the classroom factor.

Results of the ICC analysis are presented in Table 4.24. The ICCs at T1 range from 7% to 19%. The scores for the OTS were the lowest across the varieties (highlighted in yellow), indicating that classroom factors play less of a role in the variability of this construct at T1 and that variability is therefore explained to a greater extent by individual differences rather than the learning context. For each of the varieties, the ICCs were highest for CS (highlighted in blue), ranging from 15% to 19%, though these were only marginally higher than for IS, indicating that classroom factors explain 15-19% of the variance for these self-concepts.

ICCs were higher for Y/N proficiency scores (19%) than cloze test scores (10%), indicating that the classroom factor explains more of the variance in learners' vocabulary knowledge than their grammatical knowledge. This pattern is repeated at T2.

At T2, the range for the ICCs increases (.04 - .34). Again, for each variety the CS emerges with the highest ICCs; in all cases approximately a third of the variance for this self-concept is explained by classroom factors. OTS again exhibits the least variance explained by the classroom factor. All ICCs (except Classical OTS) are higher at T2, suggesting that over the academic year, the explanatory power of the classroom factor in accounting for the variability in the constructs has increased.

The proportion of variance explained by the classroom factor for MSA CS is noticeably high (34%). This is not only higher than the other varieties, but also noticeably higher than at T1, when it was half this amount (17%). This means that by T2, double the amount variability in learners' MSA CS concepts is explained by the class they are in, rather than any individual factors. This corroborates the qualitative findings that classroom features, in particular the teacher, can have a significant impact on learners' self-concept formation, which appears to grow between T1 and T2.

Table 4.24. Intra-Class Correlation Coefficients for All Self-Concept and Proficiency Test Variables at T1 & T1

		Variable	ICC	LowerCI	UpperCI	N	k (av. Group size)	Within Group Variance	Among (between) Group Variance
T1		Z YN	0.19	0.091	0.384	19	14	0.816	0.196
		Z Cloze	0.10	0.026	0.245			0.909	0.097
		MSA CS	0.17	0.075	0.358			1.312	0.274
		MSA IS	0.16	0.069	0.347			1.167	0.229
		MSA OTS	0.08	0.013	0.222		13	0.568	0.049
		D CS	0.18	0.077	0.363			1.149	0.246
		D IS	0.16	0.067	0.343			0.719	0.138
		D OTS	0.07	0.006	0.205			0.622	0.046
		CI CS	0.15	0.062	0.333			0.988	0.180
		CI IS	0.15	0.057	0.323			0.663	0.114
		CI OTS	0.09	0.019	0.236			0.553	0.054
		Current Self	0.18	0.080	0.346		13	0.848	0.166
		Ideal Self	0.16	0.068	0.253			0.920	0.054
		Ought-to Self	0.08	0.011	0.218			0.460	0.038
T2		Z YN	0.29	0.142	0.563	13	13	0.726	0.301
		Z Cloze	0.22	0.093	0.484			0.792	0.228
		MSA CS	0.34	0.169	0.610			1.116	0.566
		MSA IS	0.27	0.124	0.547			1.404	0.532
		MSA OTS	0.11	0.014	0.329		11	0.868	0.104
		D CS	0.28	0.129	0.556			1.166	0.459
		D IS	0.19	0.064	0.443			1.293	0.298
		D OTS	0.06	-0.013	0.251			0.875	0.057
		CI CS	0.27	0.123	0.545			1.131	0.428
		CI IS	0.23	0.094	0.500			1.208	0.367
		CI OTS	0.04	-0.028	0.205			0.539	0.020
		Current Self	0.33	0.162	0.601		11	0.913	0.444
		Ideal Self	0.23	0.094	0.499			1.003	0.304
		Ought-to Self	0.06	-0.012	0.255			0.589	0.039

The acceptable threshold for ICCs when it comes to disregarding the group factor is .05. Since, at both timepoints, almost all factors exhibit ICCs of >.05, this suggests that, ideally, the classroom factor should be taken into account when modelling the interactions of these constructs.

In order to investigate how classes differed from one another, as opposed to simply whether or not they did, caterpillar plots of class means with 95% confidence intervals were generated for each standardised variable in R using the lme4 package (Bates, Mächler, Bolker, et al., 2015). If the mean (0—represented by the white vertical line) is contained within the confidence interval (CI) for a particular class this means that the class average (intercept) does not deviate substantially from the grand average, i.e., the mean of all classes for that particular variable. An example of these plots is provided for the Yes/No Test variable at T1 in Figure 4.29. The full set of plots can be found in Appendix C.42.

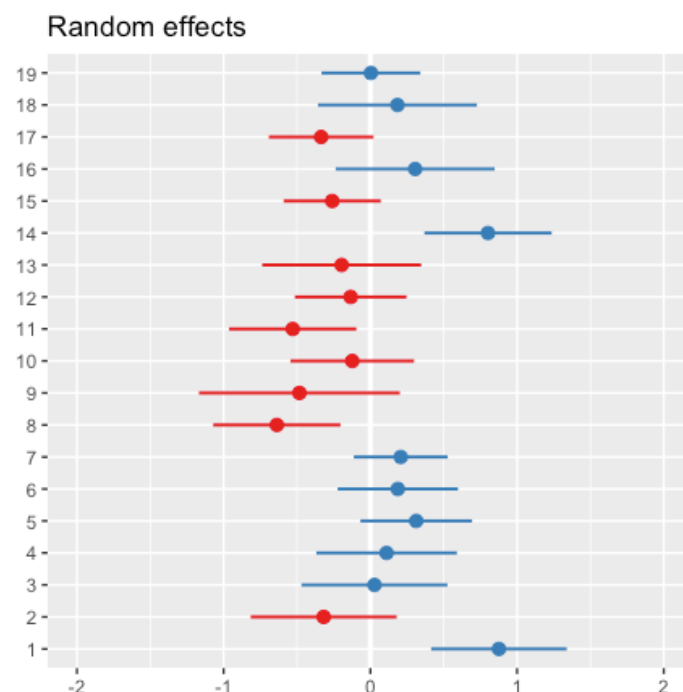


Figure 4.29. Caterpillar Plot for Y/N Test scores per classroom

From the caterpillar plot above it is clear that four school means differ from the grand mean: classes 1 and 14, whose means are above the grand average; and schools 8 and 11, whose means are below. The results from all the plots are provided in Table 4.25. (Classes with confidence intervals above 0 are written in blue, those below 0 in red).

These results offer some explanation as to the emergence of different ICC patterns. The lower the ICC score, the less likely classes are to differ from the grand mean. Very few classes differed from the grand mean when it came to OTS concepts, thus explaining the

low ICC scores for these variables. The high CS ICC scores are contrastingly explained by the high numbers of classes, almost a quarter in some cases, whose means differed substantially from the grand means.

Table 4.25. Classroom Means Diverging from Grand Mean per Self-Concept and Proficiency Test Variable				
Variable	Time	No of Classes	Variable	Class Means above/ below the Grand Mean
Proficiency Tests	T1	19	Yes/ No Tests	1; 8; 11; 14
			Cloze Tests	7
	T2	13	Yes/ No Tests	1; 8; 12; 14; 16; 17
			Cloze Tests	1; 7; 17
Overall	T1	19	Current Self	1; 8; 11; 14
			Ideal Self	7; 10
			Ought-to Self	x
	T2	13	Current Self	1; 7; 16
			Ideal Self	8; 11
			Ought-to Self	x
MSA	T1	19	Current Self	1; 8; 14; 17
			Ideal Self	5; 8; 11
			Ought-to Self	7
	T2	13	Current Self	1; 8; 14; 15; 17
			Ideal Self	5; 8; 14; 16
			Ought-to Self	7
Classical	T1	19	Current Self	1; 8; 14; 17
			Ideal Self	8; 11
			Ought-to Self	7; 14
	T2	13	Current Self	1; 8; 16
			Ideal Self	8; 17
			Ought-to Self	x
Dialect	T1	19	Current Self	1; 8; 11; 14
			Ideal Self	8; 11
			Ought-to Self	x
	T2	13	Current Self	1; 6; 7; 8; 17
			Ideal Self	8; 17
			Ought-to Self	7

Interesting to note are the specific schools which recur in the table. The means for class 1 emerge as above average for each CS concept, whereas class 8 emerges as below average for all language variety-related CS and IS variables. For OTS, in all cases where there are disparities, class 7 is either one of, or the only class to differ from the grand mean, falling below it. Class 14 also recurs frequently in the table, always with class

averages higher than 0 for a variety of CS and IS concepts. Classroom-level factors also clearly play a role in explaining differences in proficiency. Interestingly, classes 1, 7, 8 and 14 all exhibit CI which do not include the grand mean. As would be expected given the emergence of its higher CS thresholds, class 1's CI is higher than average, specifically for the Y/N tests at both timepoints and the cloze test at T2. This is also the case for class 14, with the exception of the cloze test. Class 8 emerges as below the grand average for the Y/N tests. This makes sense if proficiency drives CS and IS concepts within and between timepoints, or at the very least if they are correlated. What is striking, however, are the results for class 7, who had lower OTS scores for most varieties. The CI for class 7's cloze tests is above the grand mean at both time-points. This classroom-level pattern may offer some explanation as to the emergence of an inverse relationship between cloze tests and OTS constructs in the path models over time. However, why this relationship between lower OTS and higher cloze test scores would emerge in class 7 still requires further research.

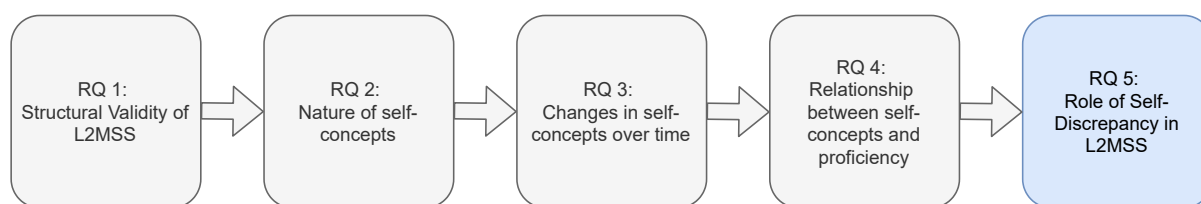
Based on the ICC results and caterpillar plots, it is clear that classroom-level factors have a role in determining learners' self-concepts and proficiency. This variable should, therefore, ideally be incorporated into the path model analysis.³³ However, none of the available options for incorporating this variable proved possible in the context of the current study. Identifying the exact role of classroom factors in motivational self dynamics in the context of Arabic and beyond thus present an obvious and promising area for future studies.

³³ In R there would be four technical options for controlling for the between-school differences:

- 1) Modelling the school-level in its own right. For this option there were too few schools to model it appropriately ($n_{jT1}=12$, $n_{jT2}=12$);
- 2) Adjusting standard errors for the clustering. This approach required having more level-2 units of analyses than the number of parameters to run. In this study, the number of parameters to estimate would have been 12 or 13 (since there were 13 classes at T2), which was insufficient;
- 3) Running the path models with an additional 12 dummy-coded school variables predicting each outcome of the model i.e., 10 dependent variables $\times$ 12=120 more paths. This would also have unreasonably increased the number of estimated parameters in the model.
- 4) Estimating a partial variance-covariance matrix, partialling out the school-effects, and using this as input data for R. Based on initial explorations of this option it was uncertain from the output generated as to whether the necessary variances had been adjusted for the school-differences or not.

4.4.5. Summary

The Factor Score Path Model analysis generated unexpected results which indicate that rather than self-concepts predicting proficiency measures over time as anticipated, the reverse is, in fact, the case. As predicted, however, different model fits and regression coefficients between variables emerged for each language variety, reinforcing the multiglossic structure of the language. Contrary to expectations, it appears that different proficiency measures predict different self-concepts. An investigation into the potential reasons for this unearthed the existence of classroom differences for multiple variables. Whilst it was not possible to integrate this factor into the factor score path models, this constitutes a necessary and interesting area for future research, particularly in the context of investigating the L2 Learning Environment, which has received scant attention in L2MSS research.



4.5. The Role of Self Discrepancy

Research Question 5:

What role does self-discrepancy play in the L2MSS?

A recurring question in more recent examinations of the L2MSS pertains to the self-discrepancy (S-D) element of the framework. This essential component of the L2MSS, based on Higgins' (1987) self-discrepancy theory, holds that motivation stems from the desire to reduce the gap between our current and future selves. To date, the vast majority of studies investigating the L2MSS have overlooked this vital component, choosing to focus instead on the relationship between the Ideal Self and outcome measures.

The analyses in previous questions have examined the nature of and relationships between self-concepts and outcome measures (proficiency measures—both subjective and objective). This section discusses the way S-D is operationalised, calculated, and measured in the current study and the exploration of its relationship to the outcome variables.

4.5.1. Calculating Self-Discrepancy

According to the L2MSS, motivational outcomes are linked to the discrepancy between the Current and Ideal selves. Figure 4.30 illustrates the hypothesised path model. There is no consensus in the limited literature on S-D in the context of the L2MSS on how to measure this construct (see Section 2.1.2.1.2.3). Higgins (1987) originally described this construct as nonideographic, one that is not supplied by participants directly. In light of this, the most straightforward conceptualisation of S-D is the difference between the CS

and IS scores. A S-D variable was thus calculated numerically by subtracting CS from IS scores, and used in subsequent path models and analyses.³⁴ The next section presents an overview of the descriptives of this variable and explores emergent differences over time and between the two sub-groups, gender and heritage background.

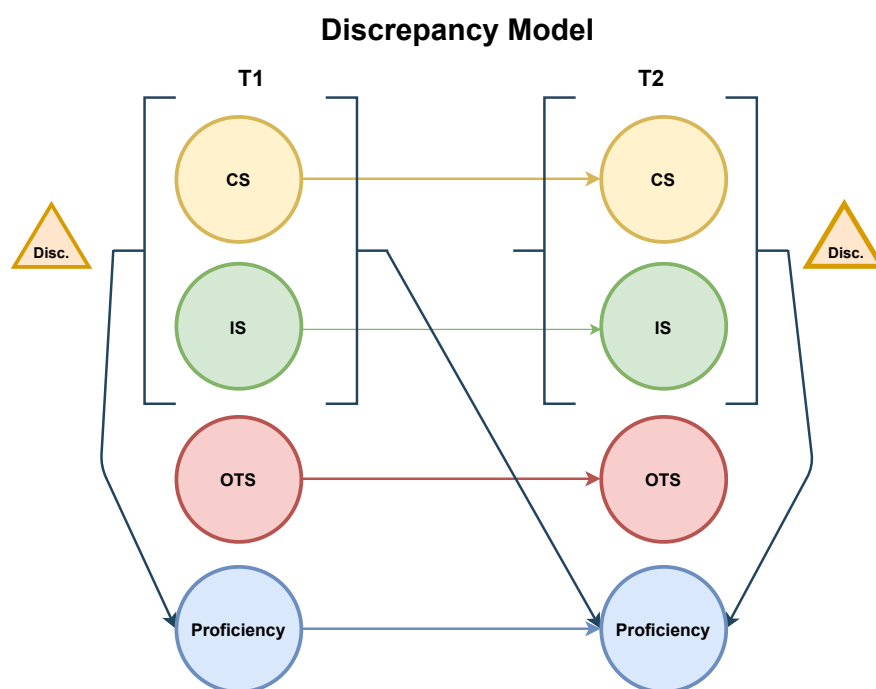


Figure 4.30. Hypothesised Self-Discrepancy Factor Score Path Model

4.5.2. Descriptive Analyses

In order to be able to conduct mean comparisons between groups and timepoints to establish whether there were any significant longitudinal or cross-sectional differences, S-D scores were calculated by subtracting the CS average-score from the IS average-scores for each language variety and for the varieties combined, labelled 'Overall'.³⁵ The descriptive statistics are presented in Appendix C.43.

³⁴ The IS Plausibility item was originally retained in order to examine whether or not this could be used as a complementary measure of S-D, given that it loaded onto both the CS and IS in the CFA. However, correlation coefficients between the Plausibility and S-D variables were either non-significant, or significant but low (see Appendix C.44), indicating that these variables did not tap into the same underlying construct. This is possibly explained by the theorised quadratic nature of S-D, as is explained later in this chapter. Given these limited correlations and the fact that plausibility was measured by only one item, further analysis of the construct was not pursued. The exploration of plausibility as a component of the L2MSS constitutes an interesting area for future research.

³⁵ Factor scores were not used to run T-tests since these values are standardised around a mean of 0, and are thus unsuitable for mean comparisons.

The largest discrepancy mean at both timepoints emerges for Dialect ($M_{T1}=1.006$; $M_{T2}=1.412$). The smallest discrepancy at T1 was Classical ($M_{T1}=.662$), whereas at T2 this was MSA ($M_{T2}=.705$), although this was only 0.17 below Classical.

Standardised values for this variable were computed and examined for outliers. Only one z-score entry emerged as outside the ± 3.29 threshold and was excluded from analysis. Skewness and kurtosis raw values and z-scores also indicate that the data met the criteria for assumptions of normality and thus parametric tests were employed.

4.5.3. Comparing Self-Discrepancy Scores

4.5.3.1. Whole Sample

4.5.3.1.1. *Between Varieties*

A paired samples t-test revealed that at T1, the discrepancies for MSA were significantly higher than the Classical, a difference of 0.22, ($SD=1.08$, 95% CI [.091, .353], $t(262)=3.341$, $p=.001$, $d=.20$).

Scores for Dialect were also significantly higher than for the Classical variety ($\Delta=-0.34$, $SD=1.00$, 95% CI [-.462, -.220], $t(263)=-5.531$, $p=.000$, $d=.34$), whereas the difference between the self-discrepancy scores for MSA and Dialect did not reach significance.

At T2, no significant difference between MSA and Classical emerged. However, the scores for Dialect were significantly higher than both MSA ($\Delta=0.71$, $SD=1.96$, 95% CI [.414, 1.000], $t(173)=4.763$, $p=.000$, $d=.36$) and Classical ($\Delta=-.690$, $SD=1.80$, 95% CI [-.959, -.421], $t(173)=-5.060$, $p=.000$, $d=.38$), indicating a larger gap between learners' CS and IS concepts in Dialect in comparison to the others.

4.5.4. Longitudinal Developments in Self-Discrepancies

With a Bonferroni adjusted p -value of .0125 required, the self-discrepancy average-score for Dialect emerged as significantly different between timepoints ($t(169)=-3.502, p=.000$), with the discrepancy increasing from .914 to 1.396.

4.5.4.1. Sub-samples

An independent samples t -test using a Bonferroni-adjusted significance level of (.05/4 tests=.0125) revealed there were no differences at either timepoint between genders when it came to self-discrepancy.

As for differences between HL and NHL learners, at T1, significant differences emerged for all language varieties, with HLs' discrepancy scores being noticeably lower than their NHL counterparts, as reported in Table 4.26. At T2, there were also significant differences for MSA and the Overall self-discrepancy scores, again with HL self-discrepancy scores much lower than the NHL scores. This indicates that at T1, across all varieties, the distance between HLS' CS and IS concepts was smaller, and at T2 this was maintained for MSA and Overall.

Table 4.26. Independent Samples T-Test—Self-Discrepancy Differences between Heritage Backgrounds

Self-Discrepancy Variable	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower CI	Upper CI
MSA Discrepancy T1	4.148	262	0.000	0.7473	0.1801	0.3926	1.102
Classical Discrepancy T1	4.296	262	0.000	0.7565	0.1761	0.4098	1.1032
Dialect Discrepancy T1	4.472	263	0.000	0.8237	0.1842	0.461	1.1864
Overall Discrepancy T1	5.01	263	0.000	0.7933	0.1583	0.4815	1.1051
MSA Discrepancy T2	3.318	172	0.001	0.781	0.2354	0.3164	1.2456
Classical Discrepancy T2	1.894	172	0.060	0.4104	0.2167	-0.017	0.8383
Dialect Discrepancy T2	0.224	172	0.823	0.0378	0.1686	-0.295	0.3706
Overall Discrepancy T2	3.302	172	0.001	0.4098	0.1241	0.1648	0.6547

4.5.5. Path Model Analyses

The analysis of the scores for self-discrepancy gives descriptive insight into the gaps between learners CS and IS concepts within timepoints and the changes in these over time. In order to examine whether or not there were any interactions between this construct and proficiency, path model analyses were employed.

For the purposes of the path model analysis, the factor scores generated by the CFA were used to calculate the difference between learners' CS and IS. This was done for each language variety in Excel by subtracting the CS factor scores from the IS ones. This new variable was labelled *CS-IS Discrepancy*. In using the factor scores, this takes into account the weighting of each item and the variances and covariances between them. Figure 4.31 shows the models for Overall³⁶. The complete output for this model can be found in Appendix C.45.

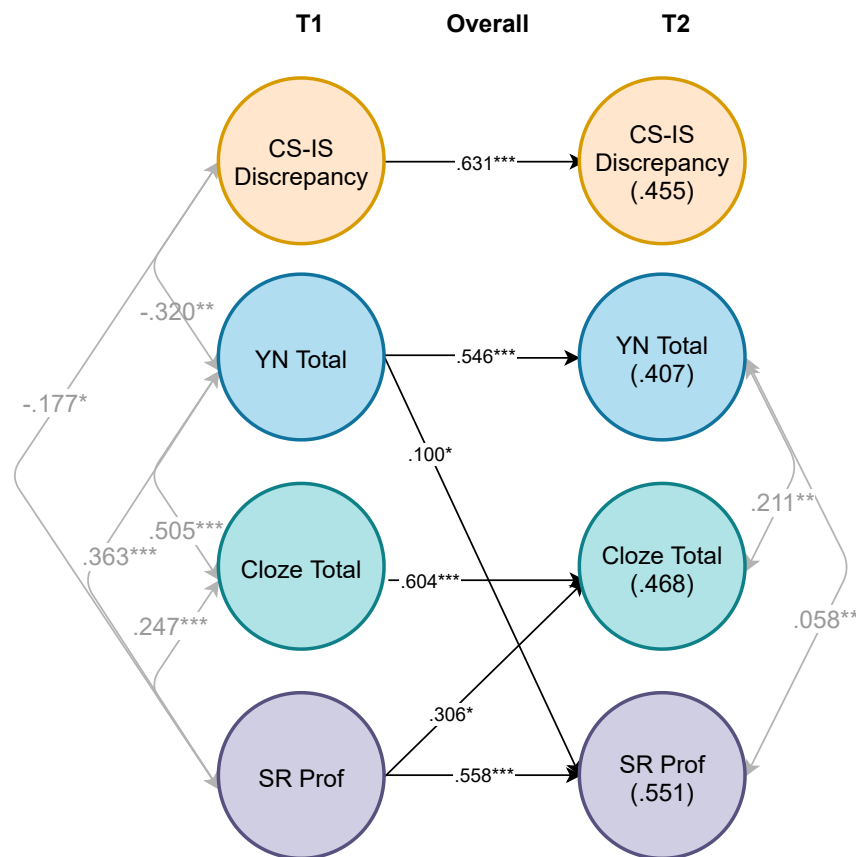


Figure 4.31. Self-Discrepancy Factor Score Path Model – Overall

³⁶ Separate models were run for each language variety. All resulting models exhibited the same patterns in relation to the Self-Discrepancy variables as the Overall model in Figure 4.31.

What is evident from these models in terms of the CS-IS Discrepancy variable is that, firstly, this variable has no predictive power when it comes to outcome measures over time. No significant pathways were recorded between this variable at T1 and any of the proficiency measures at T2.

Within timepoints, CS-IS Discrepancy shared covariances with the Y/N Proficiency measure (-.320), and the Self-Rated Proficiency scores (-.177), but not the cloze test at T1. These covariances, however, are small and negative, indicating that as the discrepancy increases, there are small decreases in the proficiency measures. At T2, it is not correlated with any of the outcome variables.

The lack of a relationship between CS-IS Discrepancy and the outcome variables at T2 and across time could be explained in two ways: firstly, the outcome measures in this study related directly to proficiency, as opposed to projected future behaviour. Moreover, the Y/N and cloze tests represent objective outcomes, rather than subjective measures of anticipated, or hoped for behaviour as represented by Intended Effort. This supports the point raised by Al-Hoorie (2018) that the constructs of the L2MSS appear to have less explanatory power when it comes to objectively measurable outcome variables, such as proficiency and achievement.

The second possible explanation for the absence of a relationship between CS-IS Discrepancy and the dependent variables (both within and across timepoints) relates to an important but overlooked characteristic of self-discrepancy, as described in the original theories. Dörnyei (2009) stipulates that in order for an IS to be motivating it should not be seen as comfortably within reach but should be plausible. The implication here is that there is an optimal level for self-discrepancy, presumably before and after which its power to motivate is less effective. In other words, where the discrepancy is small, it has little motivating power, since there is no need to change one's CS substantially in order to achieve the IS; and where the discrepancy is large, the gap may seem unfeasible, too large to overcome, and hence there is no point making much effort either. In theory, there is a range between these extremes where the gap is large enough to motivate the learner to try to narrow it, but not so large that narrowing it

seems impossible. With this conceptualisation in mind, the relationship between self-discrepancy and outcome measures is not linear but quadratic. Plotting these two variables should result in a bell-curve. A visual articulation of the theory is represented in Figure 4.32.

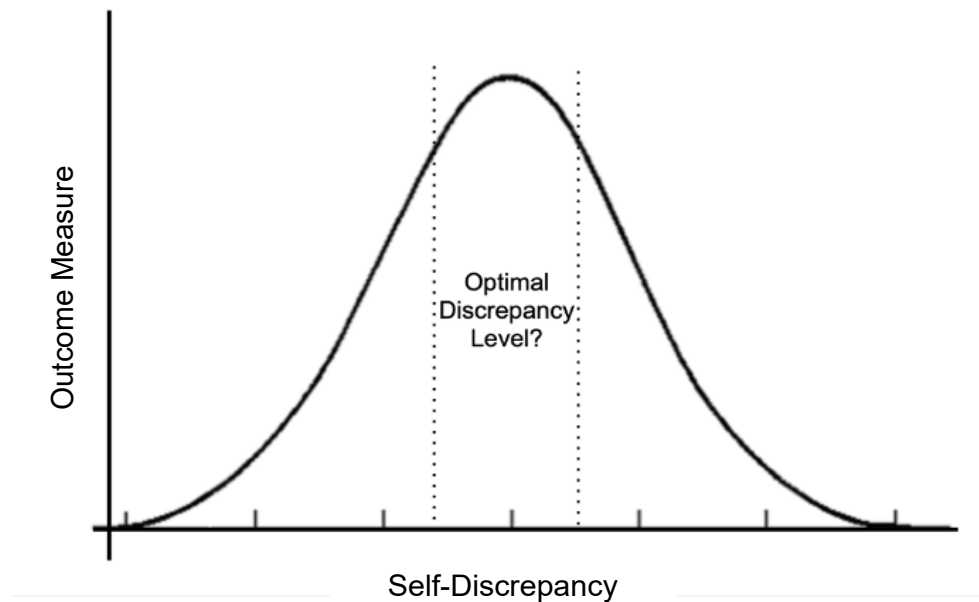


Figure 4.32. Hypothesised Optimum Discrepancy Plot

Since substantive linear effects did not emerge in the path models between self-discrepancy and the outcome variables, an exploration was undertaken of the possibility of non-linear effects. To examine whether a quadratic relationship was at play, the CS-IS Discrepancy scores were squared, creating a new set of variables which were then also modelled. No significant relationships between this new variable and any outcome measures were observed when the path models were run in R, thus the results are not reported.

CS-IS Discrepancy scores for Overall and MSA at T1 were also plotted against the proficiency test variables at T2, along with the quadratic trend lines for each. The distribution of plots was then visually examined, along with the R^2 scores for the trendlines, to determine whether a quadratic relationship could be discerned between the variables. The plots for Overall are included in Figures 4.33 and 4.34.³⁷

³⁷ Since the plots for MSA were very similar, only the Overall plots have been provided for the sake of brevity.

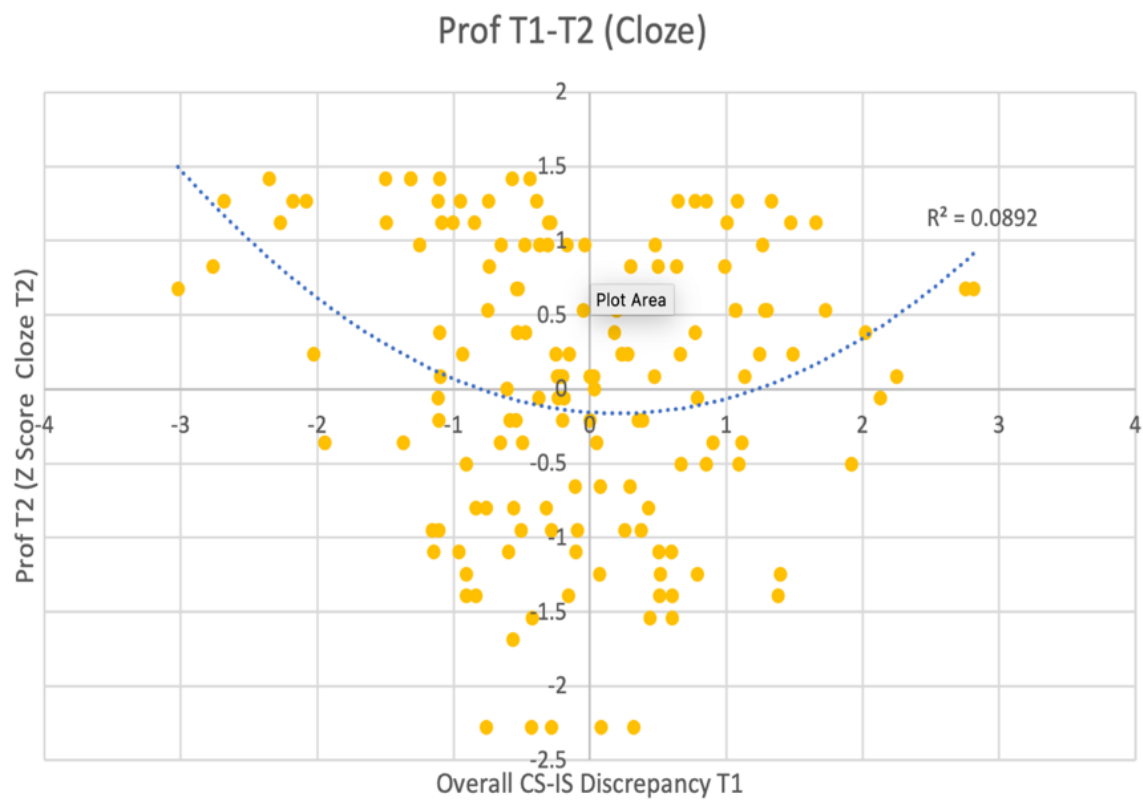


Figure 4.33. Scatter plot of Overall Self-Discrepancy at T1 against Cloze Test Scores at T2

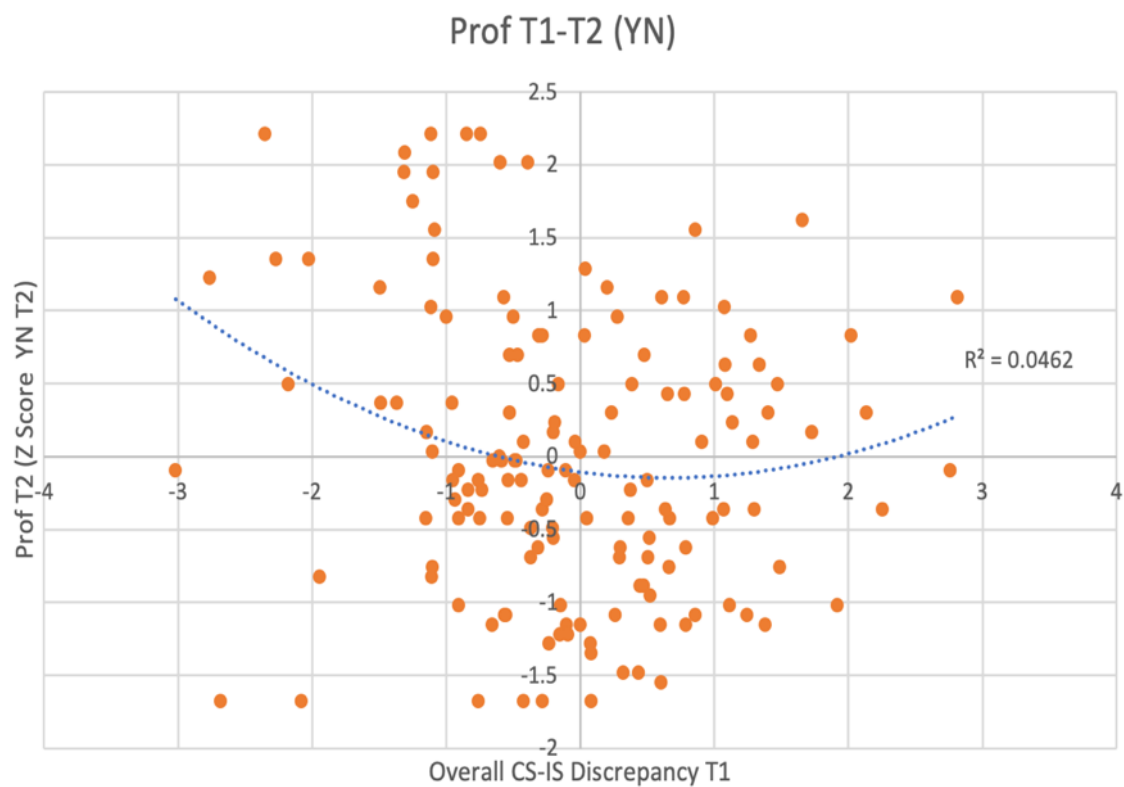


Figure 4.34. Scatter plot of Overall Self-Discrepancy at T1 against Y/N Test Scores at T2

A visual examination of these scatter plots for CS-IS Discrepancies and multiple outcome measures show that the anticipated bell-curves do not emerge, in fact where a curve is apparent, it is inverted. This means that those with discrepancy scores at the lower and higher ends of the scale scored higher on this proficiency test at the end of the year, whereas those with middling discrepancy scores scored lower. If the outcome measure is resultative of motivation, then this is the opposite of the hypothesised relationship. There is no optimal gap; rather, if anything, it is better, at least according to a visual examination of these results, to have either a very large or very small gap between one's current and IS concepts. A similar, though slightly less pronounced, pattern emerges with the Y/N test scores. The R^2 scores for these curves, provided for Overall and MSA in Table 4.27, indicate that very low percentages (<20%) of the variance in the data are explained by the quadratic trendlines.

Table 4.27. Percentage of Variance Explained—Self-Discrepancy and Proficiency Measures

	T1			T2		
	SR Prof.	Cloze	Y/N	SR Prof.	Cloze	Y/N
Overall Self-Discrepancy T1	12%	9%	19%	8%	3%	5%
Overall Self-Discrepancy T2				7%	9%	3%
MSA Self-Discrepancy T1	9%	9%	14%	5%	6%	3%
MSA Self-Discrepancy T2				4%	5%	5%

A further complicating factor in regard to the optimum discrepancy theory relates to the use of proficiency as an outcome measure. Theoretically the bell curve plots S-D against 'motivation'. Learners with small S-D scores should theoretically have the same levels of motivation, regardless of which end of the curve they are positioned. The outcome measure used in this study was proficiency, a product of motivated behaviour. Learners who have a very high CS and IS, with a small gap between them are likely to have higher levels of proficiency, related to their high CS scores. Those with very low CS and IS concepts but commensurate discrepancies would likely have low levels of proficiency. Both would theoretically be equally (de-)motivated to learn the language.

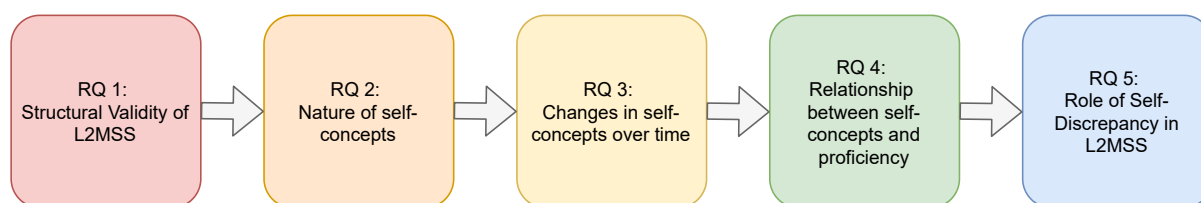
In order to capture motivation through proficiency, changes in proficiency scores between T1 and T2 were examined. These are hypothesised to capture the effort put into learning. Small improvements in proficiency may suggest small amounts of actual

effort or motivated behaviour, and large improvements more motivated behaviour. S-D was thus modelled against this new variable *Changes in Proficiency* (calculated by subtracting the T1 proficiency scores from the T2 ones). If the above hypothesis is supported, learners with smaller self-discrepancies may indeed have vastly different proficiency levels, but since their levels of motivation would be similar, the effort they put in to improving their skills should be theoretically proportionate. Thus, the gains they make in proficiency should be equal, circumventing the issue raised above pertaining to the use of proficiency as the outcome measure. This path model, involving S-D scores at T1 and T2 along with the changes in proficiency scores for both the Y/N and cloze tests, was run; however, no significant results emerged. When these variables (S-D at T1 and Changes in Proficiency) were plotted against each other, the R^2 scores indicated that less than 1% of the variance was explained.

4.5.6. Summary

Self-discrepancy appears to play no linear role in predicting proficiency variables across time. Nor does it appear that there is a significant quadratic relationship at work. This appears to call into question the role of S-D in the L2MSS model. The implications of these findings are discussed in the next chapter.

4.6. Summary of Findings



This study examined the validity and application of the L2MSS in the context of UK secondary school learners of Arabic using a combination of integrated quantitative and qualitative methods. The first research question addressed the structural validity of the L2MSS and self-rated proficiency in relation to three Arabic language varieties—MSA, Classical, and Dialect. Items for each of the three self-concepts—Current, Ideal, and Ought-to Selves—loaded onto the latent factors for each of the three varieties. These findings confirm the validity of the model thus conceptualised and offer support for the treatment of Arabic as a multiglossic language. Findings in response to the second research question were most extensive. The examination of the nature of the motivational self-concepts drew upon the survey and interview data. MSA was observed to have the lowest scores across all three self-concepts. Though IS scores for all varieties were higher than CS scores, as expected, OTS scores were lower than CS scores. HLs were also observed to have some advantages in the realms of the CS and IS, however no significant differences emerged between scores for males and females. The qualitative analysis shed light on the multiple dimensions constituted in each self-concept: the CS appears to include an affective dimension (including high levels of despondency), as well as themes of comparison and complexity; responses related to the IS raise questions regarding the operationalisation of the future, the absence of Arabic-related FS images for a large number of participants, as well as the hypothetical nature of learners' self-concepts; and finally the OTS appears to incorporate both own/other sources, each of which can be further dissected into obligation fulfilment and consequence avoidance. Furthermore, the findings suggest the necessity of incorporating the Past Self, since it informs both the CS and the IS. Research Question 3 involved an examination of whether there were any longitudinal changes in the CS, IS, and OTS or proficiency measures. Unexpectedly, there were no changes observed in the sample as a whole between T1 and T2. Although their scores do not improve except for

MSA, HLs end the year with significantly higher self-concept and proficiency scores. The relationship between self-concepts and proficiency outcome measures was next examined in response to the fourth research question. Literature on the L2MSS proposes a model in which the IS predicts the outcome measures. However, in this study the IS was not found to predict self-rated proficiency or proficiency test scores. Instead, a proficiency-driven model emerged, in which proficiency scores at T1 predict self-concepts at T2 in various ways. The original L2MSS model incorporates Higgin's (1987) self-discrepancy theory, which posits that motivation derives from the desire to reduce the gap between the CS and the FS. The data was analysed for any link between self-discrepancy and the outcome measures in response to the final research question. No evidence was found to support a relationship—either linear or quadratic—between the self-discrepancy construct and the proficiency measures, or changes in proficiency. These findings are discussed in the next chapter along with limitations, and implications for research and pedagogy.

5. DISCUSSION

The purposes of this study were several. Firstly, the structural validity of the L2MSS—revised to incorporate measures of the CS—in the context of secondary school learners of Arabic was examined. Then the study inspected the nature of participants’ language variety-specific motivational self-concepts, and the interactions of these self-concepts with each other and with proficiency-based outcome measures. A mixed-methods approach was adopted, with each component examined through multiple lenses to capture the complexities of the systems at work. Focus was also dedicated to examining longitudinal changes in the self-concept and proficiency outcome measures. One of the central, but largely overlooked, components of the L2MSS—self-discrepancy (see Section 4.5)—was also examined in an attempt to delineate its exact nature and role in the model.

In the sections that follow, each research question is addressed in turn. The major findings in relation to each are summarised. This is followed by a discussion of these findings and pertinent themes which arose during the analyses, highlighting, where relevant, implications for research and pedagogy.

5.1. Structural Validity of the L2MSS

RQ1: What is the structural validity of:

- a. the L2MSS construct as it relates to Arabic, as measured by the motivation survey scales?*
- b. the multiglossic self-rated proficiency construct, as measured by the self-rated proficiency scales?*

5.1.1. Overview of Findings

Support for the underlying, revised structure of the L2MSS—incorporating the Current, Ideal, and Ought-to selves—emerged from the responses provided by learners using the newly developed Motivation Survey scales for each of the three language varieties—MSA, Classical, and Dialect. The structural validity of the revised L2MSS was confirmed

by the results of the CFA, and items used in the survey scales met the criteria for reliability. Survey scale items, with the exception of *IS Plausibility*, loaded significantly onto the three expected latent self-concept factors for each language variety—as well as when all varieties were combined into one ‘Overall’ high-order construct model—at both timepoints. These findings offer initial support for the validity of this adapted L2MSS in the context of learners of Arabic in UK secondary schools.

The different item loadings for each language variety also provide evidence in support of the hypothesis that learners interact differently with each. Thus, while these varieties can be amalgamated into an Overall Arabic higher-order construct, to focus on this construct alone is to run the risk of overlooking specific dynamics at play in relation to each variety.

In the context of SR Proficiency, three distinct factors also emerged in the CFA corresponding to MSA, Classical, and Dialect. Two items emerged as problematic, both relating to reading, which were considered to tap into basic literacy skills. These were removed from further analysis, since the majority of students scored highly on these, most likely due to the number of years they had been studying Arabic and their exposure to and interaction with religious texts.

5.1.2. Discussion of Emergent Themes

5.1.2.1. The Problem with Plausibility

The exception to the hypothesis concerning the structural validity of the L2MSS was the Plausibility item. This was theorised to load onto the IS factor, given Dörnyei’s (2009) inclusion of it in his list of necessary IS conditions. However, including this item produced less satisfactory model fit indices for all three language varieties, and it was thus excluded from the CFA.

The plausibility item was worded as ‘*This is very likely—This is very unlikely*’ in relation to the opening statements, ‘*In the future I will be someone who can* (perform variety-

related skill)’. The problematic fit of this particular item prompted reflection concerning its exact nature and why it may not have loaded neatly onto the IS variable.

Although the item was intended to identify the likelihood of achieving a specific future state, it seems that the concept of plausibility inherently relates to both learners’ present and future images of themselves. The likelihood of an event occurring, or a state being achieved in the future depends, at least in part, on current circumstances. In this sense, the Plausibility item does not tap solely into the IS, but also into the CS as well. Accordingly, this item could be theorised as relating to the idea of self-discrepancy, or the gap between the two self-concepts. The larger the gap, the less plausible the attaining of a FS image. Counterintuitively, however, correlations with the self-discrepancy variable were not observed. This is possibly due to the theorised quadratic nature of self-discrepancy trajectories (see Section 4.5.5).

The lack of a straightforward loading of this item onto the IS variable may also be related to the wording of the item itself. The purpose behind the Plausibility item was to identify whether attaining the FS described in the opening statement was *anticipated* by students—something they are considering or pursuing in the future. However, the use of the word ‘likely’ may have inadvertently tapped into the idea of *feasibility*—whether the FS described is realistic and actually attainable given participants’ current concepts of themselves, as well as how they envision themselves in the future. It seems, based on the suggested cross-loadings of Plausibility onto the CS and IS, that learners understood this item to refer to the latter interpretation of ‘likely’—whether or not they could reasonably achieve this future self, regardless of whether or not they desired it, or deemed it possible. Al-Hoorie (2018) raises this issue of ambiguity in the wording of items, though in the context of the IS, which is usually phrased along the lines of: *I can imagine myself...*. He explains:

If a learner cannot imagine herself mastering English someday, this could additionally mean that she does not believe she can do that (self-efficacy), that she does not want to do that (value of the activity), that she experiences a complete absence of motivation (amotivation), that she does not need to do

that (e.g., she has already mastered English), or any other interpretations different learners might conjure up. (p.736)

An interesting area for future research would be to unpick these distinctions through separate items relating to anticipation and feasibility, and to examine their relationship with different outcome measures.

5.1.2.2. Arabic or Arabics?

The second hypothesis regarding the structural validity of the L2MSS in the context of Arabic relates to the distinctiveness of the three language varieties: MSA, Classical, and Dialect. The fact that responses to all three varieties could be combined and produced the same factor structure indicates that there is an overall higher-order construct of 'Arabic' which can be examined. The majority of items loaded as expected onto the hypothesised underlying variables: MSA, Classical, and Dialect. Importantly, the emergence of distinct factor loadings for each variety also indicates the necessity of examining these separately and lends validity to the claim that the multiglossic nature of Arabic is not just theoretical but is supported by the finding that learners interact differently, in terms of motivation, in the context of each. Conflating these varieties in a research context runs the risk of overlooking important dynamics in learners' interactions with the language.

References to different varieties of Arabic in the interview data also provide evidence for the ecological validity of this taxonomy of varieties. Learners spoke of 'worldly' Arabic which they saw as a 'general' version of the language, aligning with what they studied in their GCSE classrooms, which was MSA; 'mosque' Arabic which referred to the content covered in extra-curricular, religiously-oriented classes focused on the comprehension of religious texts and therefore denoting the Classical variety; and colloquial Arabic, though references to this were least frequent and mainly made by HLs, indicating a general lack of awareness of the distinction between this variety and MSA.

5.1.2.3. 'I can already read Arabic'

Notable exceptions to the SR Proficiency scale item loadings included *Word Reading with Diacritics* and *Word Reading* (without Diacritics). The former was hypothesised to load onto the Classical variety, since this skill is confined, in authentic contexts of use, to reading the Quranic script. Given the large skewness values for this item, and the fact that it did not load strongly onto either Classical or even Modern Standard Arabic an alternative approach was adopted. The two were removed from the CFA and assumed to be indicators of a separate Basic Literacy Skills variable which, in the context of this population, was redundant given that the vast majority of participants had been learning the language for a number of years and scored highly on these scale items.

In the interviews, many participants asserted that they could already 'read Arabic'. What the SR Proficiency scores and follow-up questions in the interviews regarding these statements make clear is the conflation of skills subsumed under the umbrella of 'reading'. Given the shallow orthography of the Arabic script (when written with diacritics), the ability to phonetically decode the language is often acquired quickly (Abadzi & Martelli, 2014). This explains the high mean scores for this item in the SR Proficiency scales. However, phonetic decoding does not entail any form of comprehension. Thus, when asking whether or not students can 'read' a text, it is imperative that they are also asked to what extent they can understand what they are reading. While the items pertaining to phonetic decoding may have been redundant in the context of this study, they may prove useful as a Basic Literacy Skills variable in other contexts involving early *ab initio* learners of the language.

5.1.3. Summary

Thus far, the tripartite structure of the L2MSS model appears to be supported in the context of Arabic secondary school learners, since the three motivational self-concepts (Current, Ideal, and Ought-to Selves) emerge as separate, though related constructs. The instruments developed provide reliable tools for capturing the self-concepts and SR Proficiency levels. Most importantly, they confirm the clustering of skills related to different language varieties, endorsing the need for measures of motivation and self-

rated proficiency that capture the multiglossic nature of Arabic, rather than treating the language as a monolithic entity.

Taken together, the findings regarding the structural validity and the in-depth interviews contribute to the field in a number of ways: First, the integration of the CS along with the two future self-components was confirmed. In addition, this tripartite structure of the motivational self-concepts was confirmed in three Arabic language domains, while previous studies have focussed on motivational self-concepts in relation to monoglossic languages, the most predominant of which is English (Boo et al., 2015). Furthermore, the in-depth follow-up interviews provided learner statements confirming the tripartite nature of motivational self-concepts in three domains, that converged with the quantitative findings. This goes beyond previous research in which *either* interviews *or* surveys have generally been used.

5.2. The Nature of Motivational Self-Concepts

RQ 2: What is the nature of Arabic learners' motivational self-concepts? How do learners' motivational self-concepts vary based on:

- a. Arabic language variety;*
- b. Gender;*
- c. Heritage Learner status?*

Findings in response to this research question highlighted the necessity of expanding the breadth and depth of detail involved in each self-concept. Given the extensive empirical findings in response to this question, discussion of the nature of Arabic learners' self-concepts follows a thematic approach, taking each self-concept in turn, presenting first the main findings pertaining to them individually, before discussing the complexities, implications and limitations involved.

5.2.1. Current Self

The CS is a relatively unexplored component of the L2MSS. Few studies (see, e.g., Thorsen et al., 2020) have incorporated it into the study of motivation despite its

central theorised nature in calculating self-discrepancy—an integral mechanism of the model which purports that motivation is derived from the desire to reduce the gap between an individual's current and future self-concepts. The CS was measured through the use of *Self-perception*, *Ability*, and *Confidence* items in the Motivation Survey scales. The CS mean scores emerged as lower than the IS for each variety of the language, but higher than the OTS at both timepoints for the whole sample. CS scores for MSA were significantly lower at T1 than Dialect and Classical Arabic, with the scores for Classical being the highest. At T2, the scores for MSA CS were also significantly lower than those of Dialect and Classical, though there was no significant difference between these two, caused by a decrease in both sets of scores resulting in them being within closer range of each other. Unexpectedly, no significant differences between genders emerged at either timepoint. There were also no significant differences between HLs and NHLs at T1 for the CS for any of the varieties. By T2, however, HLs had significantly higher CS concepts across all varieties. Analysis of the interview data revealed further layers of complexity to the CS: the necessity of incorporating an affective dimension; the intrinsic role of comparison in determining the CS; the complexity of learners' self-assessments; and the need to disentangle the language from the course of its study. Each of these are discussed in turn in the sections below.

5.2.1.1. Explaining the Variety Hierarchy

It is perhaps not surprising that participants' scores for Classical Arabic are highest given their predominantly Muslim backgrounds, as a result of which they engage in reading the Quran. As attested to by the high *Word Reading with Diacritics* item scores on the SR Proficiency scales and their frequent mention of this skill (and the practice of it) in interviews, participants' ability to phonetically decode is well established. The finding that the scores for Dialect are higher than those for MSA at T1 for the entire sample, as well as amongst both HLs and NHLs individually, requires explanation. For HLs to have higher levels of self-concept in relation to Dialect is expected; however, for this to be the case with NHLs is puzzling for two reasons: firstly, it implies a level of proficiency in colloquial Arabic which these learners do not have (nor do they profess to having such proficiency in the SR Proficiency items pertaining to Dialect); and secondly,

because it implies that they have had opportunities to communicate with native speakers, allowing them to assess their level of spoken skills.

One possible explanation for this apparent paradox relates to the opening statement for Dialect, which reads: *'In the future, I will be someone who can communicate fluently with native speakers'* in response to which learners were asked to rate their current abilities in this domain. Unlike in the SR Proficiency scales, 'dialect' here is not specifically mentioned. It may be the case that learners consider themselves able to hold conversations with native speakers (NSs) in the standard variety (MSA), thus making this statement currently true of them, to some degree. The opportunities to speak to NSs likely relates to either their classmates, since there were HL students in most classrooms, or their teachers, many of whom were from Arabic-speaking backgrounds. With these individuals involved directly in the learning process, it would be expected that communication would take place in MSA, the language of the GCSE. Limited communication on holidays abroad in Arab countries may have also provided the context for some learners' communication. However, of the NHL students who mentioned speaking abroad, very few of them made mention of the colloquial varieties being spoken. This suggests a lack of awareness amongst learners of the integral nature of dialects to authentic spoken settings. Whilst it is indeed possible to communicate in MSA with Arabic speakers, there is a general consensus that in order to communicate 'fluently' as the opening statement says, knowledge of a colloquial variety(s) is vital (Trentman & Shiri, 2020). The question is therefore raised as to the legitimacy of an approach to teaching Arabic which does not at least make learners aware of the integral nature of Arabic dialects to everyday communication.

An alternative approach to explaining this emergent hierarchy of varieties in the context of the CS is to ask not why scores for Dialect are higher than MSA, but why the scores for MSA are so low, particularly given that this is the variety learners have been exposed to throughout their schooling. One possible explanation refers to the wording of the opening statement for MSA: *'In the future, I will be someone who can read books, magazines, and newspapers.'* This presents the most prominent context in which MSA is used—modern written texts. However, learners' abilities in reading these authentic resources are hindered by two factors. The first is the lack of vowels in these written

texts, which makes comprehension immediately far less accessible (Eviatar & Ibrahim, 2014) than if words are spoken aloud. Given that learners appear to be unaware that day-to-day communication does not take place in MSA, they may feel it is easier to comprehend aural input than written unvowelled input, particularly if the majority of their exposure to spoken Arabic takes place within the classroom. The second issue is the accessibility of authentic materials from such sources as print journalism that are of appropriate difficulty level. Excerpts in the GCSE specification are rarely derived from authentic sources (Ramezanzadeh, in preparation), possibly given the difficulty they present to students lexically as well as grammatically. As the average-scores for the Y/N vocabulary test were low even for vocabulary pertaining to participants' GCSEs, it is therefore unsurprising that learners CS concepts are low for this particular variety.

5.2.1.2. Gender Equality

The lack of significant differences between genders is, at first, unexpected given previous studies indicating that females are more motivated to learn languages than males (e.g., Carreira, 2011; Henry, 2010). However, such studies most often focus on commonly taught languages, such as French and German. Arabic is distinct from these languages on two accounts. Firstly, it is not commonly taught in schools and, thus, does not have the associations that other MFL enjoy or suffer from. Secondly, the language has unique ties to the Islamic faith, which may contribute to the treatment of Arabic as a more 'familiar' subject, to use the words of one of the participants, a perception held equally by male and female learners. Of course, further studies into gender differences are necessary, particularly given that girls and boys are clustered within classrooms, since the majority of schools in the present sample are single-sex, and there was evidence of classroom-level effects (see Section 4.4.5.2).

5.2.1.3. Heritage Advantage

When it comes to distinctions between HLs and NHLs, the absence of any significant differences at T1 is unexpected. Initial hypotheses held that, given their exposure to various forms and sources of input outside of school, HLs would have higher CS concepts in relation to each variety, or, at the very least, the Dialect items. However, at

T1, no significant differences emerged between the two groups. One possible explanation draws upon the fact that the scale items are subjective measures. Regardless of how skilled learners actually are, these items tap into how they perceive their skills. HLs may be more aware of the reality of authentic language use in multiple contexts, and of just how far they need to progress in order to attain specific FS images, in contrast to NHLs who may have a tempered view given their lack of exposure. There is also some evidence that familiarity with colloquial spoken Arabic affords limited proficiency in accessing standard or classical written texts (Faraj, 2020), making the equal footing on which HLs and NHLs appear to start rather less surprising in the context of their MSA and Classical CS concepts.

An area of concern from a pedagogical perspective, however, is the emergence of significant differences between the two heritage background groups. By T2, HLs' CS scores were higher than their NHL counterparts across all three varieties, indicating that they had grown faster in their self-perceptions, confidence and ability levels. This supports the findings of Husseinali (2012) who found that HLs perceived themselves as more likely to achieve speaking proficiency than NHLs, as well as the claim (thus far predominantly anecdotal) that HLs enjoy an advantage when it comes to the GCSE. This difference appears to have been precipitated both by an increase in HL scores and a drop in NHL scores for each variety. (Longitudinal changes are discussed in Section 5.3.1).

5.2.1.4. Complexities of the Current Self

While analysis of the survey items offers an insight into the strengths of learners' CS concepts as measured by the self-concept scales, the interview data shed light on the internal workings of this self-concept. Participants shared thoughts on the *a priori* components of the CS concept included in the survey (self-perception, ability, and confidence) but they also raised themes and issues which went beyond these three items. Findings highlight important areas for incorporation into the operationalisation of the CS: the basis of self-assessments; perceptions of the language; contextual comparison; the affective dimension; and complexity, each of which is discussed below.

5.2.1.4.1. *The Basis of Self-Assessments: Skill- and Grade-based*

When it came to learners' assessments of their abilities, many participants voluntarily expressed these in terms of specific skills, rather than simply providing an overall assessment of their ability in Arabic. This supports the separation of the language into domains, rather than simply an aggregated skills appraisal. Another issue with ratings of ability regards the foundations on which they are made. In this study, many learners referenced their GCSE grades when assessing their skill levels, though it is not clear whether all participants derived their self-perceptions from their current grades (reflecting the level they are working at presently) or predicted grades (those they are expected to achieve in the GCSE exams at the end of Year 11). This makes comparison within the cohort, as well as across studies, problematic, and speaks to a wider issue with regards to self-reported measures (Demetriou, Uzun Ozer & Essau, 2015). Upon which basis is the self-report being made, what is influencing this decision, and are all learners using the same terms of reference? Follow-up questions in surveys and interviews in future studies may help researchers identify and investigate the sources of such self-assessments.

Similar issues affect learners' statements concerning their confidence, which was often presented as both skill-contingent and influenced by the grading system for the GCSEs and their ranking within it. Confidence levels amongst learners were generally low, corroborating earlier findings by the researcher (Ramezanzadeh, 2015). One reason for low levels of confidence was identified to be the new 9-1 GCSE specification which, according to analysis by the researcher (Ramezanzadeh, in preparation) and multiple anecdotal reports by teachers,³⁸ is significantly harder than the previous version. This relates to two further emergent codes: *conflation* and *comparison*, discussed below.

5.2.1.4.2. *Perceptions of the Language and Conflation of Referents - Language vs. Language Course*

Learners were asked about their experience learning Arabic in order to ascertain information about their CS concepts. A frequently referenced term in relation to this

³⁸ These were collected during the researcher's communication with teachers during visits to schools for data collection purposes.

question was ‘difficulty’. The perception of Arabic as ‘difficult’ is a pervasive one (see Section 2.2.1.1.1) and has been found in other Arabic learning contexts (e.g., Aladdin, 2010). Where the source of this difficulty was clarified by learners, they mentioned a number of factors including the grammatical features of language, as well as the volume of vocabulary required for the course. Upon closer examination of the data, it appears that learners have, in this context and in others, use the term ‘Arabic’ to mean either the language itself, the specific course of study, i.e., the Arabic GCSE, or both. The sentiment ‘Arabic is difficult’ could refer to either the language or the actual course. Significant statistical differences between learners’ feelings towards Arabic (the language) and the GCSE confirm that learners’ interactions with each, at least on an affective level, differ. Additionally, a number of learners mentioned the desire to study Arabic outside of school or their experiences with extra-curricular Arabic, such as the ‘Alimah course. In relation to the latter, students spoke very highly, in contrast to their evaluations of their school experiences. It appears, therefore, that motivation towards the language itself can be distinct from motivation towards a particular course of study. Future research needs to avoid this conflation in order to determine: (a) learners’ perceptions of the language as distinct from the course of study, and (b) their self-assessments with respect to each one.

5.2.1.4.3. *Contextual Comparisons*

An analysis of learners’ perceptions is not only important to identify the referent of their self-assessment, but also to provide context for them. Descriptions of the language as ‘difficult’, for example, imply experience with or at least perceptions of other subjects which are easy (or easier). In fact, some students expressed the desire to study a European language, or Urdu with which they were familiar, since these would have been less challenging for them. This raises the theme of *comparison*, which emerged from the interview data. An individual’s CS is unavoidably synchronously and diachronously relative. In other words, the formation of a learner’s CS is influenced by both their past and current context and learning experiences and should be assessed in relation to both. This could be interpreted as a version of Ushioda’s (2009) ‘person-in-context’ approach. Rather than viewing the language learner through the prism of their social and cultural interactions, the findings suggest adjusting this lens to incorporate

their perceptions of and interactions with their personal learning contexts. In the interviews, learners made multiple comparisons between Arabic and other subjects, and between their current and past experiences of learning the language. Though it is not an easy task to determine the contextual factors against which learners are comparing and deriving their self-assessments, without such information researchers run the risk of not comparing like-with-like when it comes to learners' CS concepts. A learner expressing a lack of confidence may do so due to changes in the perceived course difficulty, despite still achieving higher grades than a learner with similar low confidence levels due to a perception of themselves as a poor language learner. This oversimplification of constructs, one may argue, is inherent within quantitative research. One solution, in line with a CDST approach through which many of the findings of this study can be viewed, is to extend the net of measurement so as to capture as much of this detail as possible before deriving conclusions. This study has gone some way in capturing the intricacies of participants' interactions with their learning contexts, supporting the view that 'contextual factors will now need to be understood as key dimensions of the [language motivation] system itself' (Al-Hoorie, Hiver, Kim et al., 2021, p.9).

5.2.1.4.4. *Affect: How does that make you feel?*

Affective statements also provide context for learners' self-assessments and appear to constitute an important component of learners' CS concepts. Previous studies have reported substantial relationships between learners' emotions and their achievement (see, for example: MacIntyre & Gregersen, 2012; Pekrun, 2006; Pekrun, Lichtenfeld, Marsh, Murayama & Goetz, 2017).

To date, however, limited reference has been made to emotions in the L2MSS literature. Some studies have investigated the links between emotions and components of the L2MSS, predominantly the IS (e.g., Saito et al., 2018). However, to date, affect has not been integrated into the L2MSS. This admission was made by Dörnyei himself, which he links to the model's cognitive-oriented origins (Dörnyei & Ryan, 2015). This 'emotional deficit' (p.10) is surprising given that: (a) the model attempts to capture learners' hopes and desires, both domains with significant affective links; and (b) links between

emotion and cognition have been described as ‘intimately intertwined’ neurologically (Schumann, 1997, p.238), and were in fact included in by Markus and Nurius in their initial 1986 paper. Empirical work in other educational contexts has demonstrated links between achievement and emotions, adding strength to the argument of including them in the L2MSS CS concepts (e.g., Pekrun et al., 2017). Positive emotions, such as the enjoyment of learning, are considered to ‘preserve cognitive resources and focus attention on the learning task, support interest and intrinsic motivation, and facilitate deep learning’ amongst younger learners, whereas negative emotions ‘reduce cognitive resources and task-related attention, to undermine both intrinsic and extrinsic motivation, and to promote shallow information processing’ (p.1655).

Participants in the current study reported both positive and negative feelings in relation to learning Arabic. One particularly noticeable finding was the high levels of despondency expressed by participants, often voiced in relation to the difficulty of the subject and the course content. Since emotions were not included in CS items in the current study, it is difficult to ascertain the impact of these feelings upon actual achievement. However, the prevalence of these affective states, largely in response to the course, does raises an ethical question: If a large number of students are experiencing such states as despondency in relation to a language course, how appropriate is the design or delivery of that course of study?

Not all learners expressed negative feelings towards their Arabic learning, however; a number spoke of enjoying the subject and of finding it interesting. An investigation of the clustering of these feelings revealed strong underlying classroom trends, highlighting the impact of the learning environment, in this case the teacher and teaching style, on learners’ affective positions towards Arabic. It could be argued that feelings constitute part of the learning experience (L2LE), rather than the CS, since the former has been conceptualised to include ‘attitudes and evaluations of the present learning environment’ (Al-Hoorie, 2018, p.724). The very ambiguity of where this affective component falls touches upon the issue of how to operationalise the under-theorised L2LE (Dörnyei, 2019a). Nevertheless, emotions constitute an important layer of participants’ interactions with the language and offer a way to expand the conceptualisation of learners as a system in which there are multiple dynamics at play,

all of which influence and are influenced by motivation. An important point to note, too, from a pedagogical perspective, is that negative emotions cannot and, in fact, do not necessarily have to be avoided. The processing of difficult emotions, through a combination of cognition and affect, a process coined by Vygotsky (1971) as *perezhivanie*, can promote progress in language learning if harnessed appropriately. According to Lantolf and Swain (2020), 'Development emerge[s] precisely from the catharsis arising from confronting and overcoming threatening emotional circumstances' (p.103). If a perezhivaniyan approach is adopted—where teachers are aware of and engage with the emotions of their students and encourage them to think about their affective experiences of the language in the classroom—it may be possible to transform the energy generated from these feelings into a driving force towards achievement, rather than away from language learning altogether.

5.2.1.4.5. *Complexity: 'I like Arabic but...'*

Presenting emotions, and even perceptions, as binary stances adopted by participants, however, would be misleading, and an oversimplification, in many cases, of the reality of learners' positions towards the language. One prominent theme to emerge from the data was that of *complexity*. The use of the word 'but' and its synonyms when expressing views about Arabic highlighted the notion that very often learners do not have a solidified, immediately identifiable position with regards to the language in question. Rather, they have multiple voices which they may or may not try to reconcile and which can be both skill-, context-, and even time-dependent. The use of the word 'but' clearly highlights the juxtaposition of such voices, which quantitative measures would find hard to capture, let alone reconcile, if indeed they can be. This again highlights the necessity of a confluence of methods to capture the complexity of the constructs at play.

There are multiple possible explanations for the use of such nuanced phrases by participants. One is that there are components of the language that learners do enjoy, but others that they do not. Another is that they are interested in the language but struggle with the content and are, therefore, not confident. In addition to mismatches in their CS concepts, there is also the issue of the researcher effect. As a visibly identifiable

Muslim woman interviewing students in predominantly Muslim schools, the possibility of the provision of potentially 'face-saving' answers is high. Though there is no judgement on the part of the researcher, a fact that was conveyed to participants verbally, the power dynamics at play in an interview, as well as in a wider school context, where expectations as to achievement in Islamic-related subjects are both implicit and explicit, may have resulted in learners qualifying their 'negative' statements with 'positive' ones. This would serve to avoid open criticism of a language so closely tied to their religion. The researcher effect can only be alleviated to a degree (Miles et al., 2013), and there is no way to identify for sure which statements align with which explanation. What is important is for this to be borne in mind and for comparisons to be made between repeated studies conducted by researchers of differing backgrounds and visible identities so as to examine the extent of the researcher effect in this particular domain.

5.2.1.5. Summary

Pursuant to a CDST approach, which advocates analysis of the systems under scrutiny in as much detail as possible, the emergent themes hitherto outlined expand the focus when it comes to learners' CSs. Not only do they capture the affective dimension of learners' interactions with the language, but also their perceptions of the nature of the language which, reciprocally, appear to inform their interaction with it and subsequent levels of confidence. Furthermore, the broader context of learners' experiences is taken into account. Learning does not take place in isolation. Learners have both a learning history and a present, both of which appear to inform their CS concept through the act of comparison, either between their past and present experiences, or between the multiple subjects they are studying. How confident a learner claims to be in a subject will inevitably be influenced by how confident they feel in relation to other subjects, as well as how confident they used to feel in the subject in question. Isolating a learners' self-ratings of their ability and confidence misses crucial components of the system within which they are operating. Aggregated scores for CS concepts certainly have their utility and allow researchers and practitioners to quickly capture a snapshot of an internal environment. However, this runs the risk of overlooking levels of complexity both within and between learners' CSs as demonstrated in the interviews through the

frequent pronouncement of hedged statements by students when offering their views on Arabic learning. The 'But' phenomenon represents variance within learners with regards to their assessments and perceptions of the language, which also cannot be captured through aggregated item scores.

5.2.2. Ideal Self

The IS was measured via survey scale items related to desire, frequency, and vividness of the self-concepts described in the three opening statements. The plausibility item did not load as expected onto the latent IS factor and was removed from the analyses. Average-scores for the IS were higher than both the CS and the OTS across the sample. At T1, IS Dialect scores emerged as higher than IS Classical scores, which was higher than IS MSA scores. At T2, MSA was significantly lower than both Dialect and Classical, which were now not significantly different. When the sample was split by HL background, no significant differences between the two groups emerged in the context of IS at T1. By T2, this had changed, and HL scores were significantly higher for both MSA and Classical. Interview data revealed that most the frequently cited IS image was *Travel*, followed by *Religion* and *General Skill*. It also cast doubt on the reliability of learners' professed IS images, given their frequent use of hypothetical phraseology and raised issues concerning the timescale used to conceptualise the future.

5.2.2.1. Explaining the Unpopularity of MSA

That IS scores were higher than CS scores is expected. What is challenging to explain are the low MSA scores at both timepoints, given that learners are studying this variety in their classrooms. As mentioned above, a potential explanation for this is found in the wording of the opening statement which refers to the reading of '*books, newspapers and magazines*'. Whilst this is the most common domain of authentic MSA usage, it appears that it is unappealing to students. Participants appear far more interested in communicating with native speakers and/or understanding religious texts. This supports the perception of a mismatch between learners' desires, their expectations of the remit of the course content, and the actual application of MSA. In light of the findings of Donitsa-Schmidt et al., (2004) who found that school-level students of

spoken Arabic were more motivated to learn the language, the focus on MSA may be posited to underlie the static nature of learners' self-concepts and by extension their motivation. This is not to say that the learning of MSA cannot be useful in achieving goals other than reading modern texts; there is a degree of overlap between MSA and colloquial and classical varieties (Al-Batal, 2017). However, there is little evidence that learners are being supported in translating their MSA skills into realms of classical and colloquial use. Nakamura (2019) proposes the idea of domain-specific conceptualisations of motivation. The emergence of distinct scores for each language variety, themselves centred on distinct domains of usage, appears to support this approach, though not the idea of scrapping motivational selves altogether, as Nakamura goes on to suggest.

5.2.2.2. Heritage advantage

HLs, once again, witnessed developments in their self-concepts by the end of the year, at which point their MSA and Classical IS scores were significantly higher than NHLs'. In other words, at T2, their perceptions of themselves as exhibiting these skills in the future is stronger than those from non-heritage backgrounds. This is potentially explained by the finding that their CS scores are also higher at T2 and their stronger self-assessments developing the strength of their IS concepts. What is noteworthy, however, is that this does not apply to Dialect. This would be the domain where, intuitively, HLs would be expected to be stronger at both timepoints, given their connections to NSs and Arabic-speaking environments. This observation contradicts findings by Olsen (2018) who found that HLs of Maori scored significantly higher on measures of the IS. A possible explanation for the absence of significant differences in the Dialect IS scores lies in the multiglossic nature of Arabic and differing perceptions of what it means to communicate fluently in Arabic with NSs and the awareness of what constitutes this. HLs, it can be assumed, are more likely to have a grasp of the reality of authentic use given their exposure, either past or present, to Arabic-speaking individuals or communities. Few studies have examined differences between HLs and NHLs in the context of the L2MSS (Olsen, 2018), making this an area ripe for future research.

5.2.2.3. Categories of IS concepts

Analysis of the interview data revealed a number of broad categories from responses to questions about learners' Arabic-specific IS images. Travel was the most oft-cited, corroborating the findings of other studies (Graham et al., 2016). This was followed by religion, which is expected given the context within which most participants were studying and given their religious backgrounds. General references to having Arabic as a skill were the third most common, followed by the absence of any Arabic FS image, which was expressed by over a quarter of participants and appeared to be clustered in some classrooms. Job and family then followed in terms of frequency. This ordering of IS concepts aligns with the findings of the survey data given the overlap between (a) Travel and Dialect, (b) Religion and Classical, and (c) General Skill and potential Job and MSA.

These broad categories, which help to some degree in identifying which varieties of Arabic would be useful for learners, mask a complexity which deserves analysis. When it came to Travel and Jobs, two important threads were detected in the interview data: (1) the mention of FS images was very rarely tied to any awareness of the limited utility of MSA in authentic Arabic-speaking contexts, which has been aforementioned; and (2) the popularity of these IS concepts, when interrogated, masks a tendency to provide generic and/or hypothetical answers to questions concerning FS images.

5.2.2.3.1. 'If'

The *in vivo* coded theme of 'if' covered three types of IS image: the hypothetical or generic; the uncertain; and the contingent. The first relates to IS images that are framed by words such as 'only if', e.g., only if I go abroad. Such statements would usually be categorised under the umbrella of travel-oriented IS concepts. However, upon closer inspection of these responses, it appears that participants often provide these answers without IS concepts actually constituting a part of their future selves. These responses appear to be generic statements which students are able to provide if prompted, and are often undermined by further questioning, in response to which participants reveal the absence of likelihood of these statements being fulfilled. This lends support to the

integration of multiple items pertaining to the IS in any quantitative instruments, since learners may be inclined to answer positively in response to questions regarding the ability to imagine a FS which may actually constitute no part of their future self. One answer as to why learners would give generic answers is perhaps that they were caught off-guard by the questions concerning their future selves—and, by implication, the reasons they were learning Arabic—and felt obliged to say something more than just ‘I don’t know’. Dörnyei and Al-Hoorie (2017) describe Global English worldwide as ‘the default or ‘unmarked’ option of language choice’, a characteristic feature of learning the language is ‘that it does not require any special justification’ (p.462). In the context of Muslim schools, in the UK and presumably beyond, this may also be the case for Arabic. Such generic answers may represent an engagement with the language as little more than an obligatory, default requirement, in line with the ‘functional’ approach adolescent Arabic learners in Calafato and Tang’s (2019) study adopted.

This also applies to the second category, the less certain or tempered IS images. These are often couched in terms of indecision, such as ‘maybe’ or ‘might’. These are distinct from the generic ‘lip-service’ statements of the first category, since they are not undermined by the students themselves. However, such statements demonstrate a similar lack of full integration of these images into learners’ own IS concepts.

The final category, contingency, is perhaps the most noteworthy, since it reaches to the heart of the underlying concept of the IS: what constitutes the future? Learners with responses in this category would express IS images that are contingent on other future events, such as doing well in exams. Such events may or may not be related to Arabic. When a learner is asked about their Arabic-specific FS image, the question is then, when is the future? What is being referred to? And what is required to get there? Learners may have multiple Arabic-specific IS images, each of which is contingent on earlier ones. Alternatively, a learner may have one long-term goal which is contingent on multiple non-Arabic related goals. At present, this perspective on the future is not factored into the L2MSS. One suggestion for future studies would be to ask participants to outline their own IS image, what steps they would need to take in order to realise it, and the likelihood of being successful in each of those. This would be consistent with a CDST approach to examining learners’ self-concepts, since it takes into account the

importance of temporal frameworks (Elman, 2003) and the interrelatedness of the multiple factors upon which a FS image is based (Dörnyei, 2017). The problem with this approach, however, is where to stop, since it is likely that each of those pre-FS steps, which themselves constitute FS images, are likely also contingent on a number of other steps, and so *ad infinitum*. Determining the likelihood and genuineness of each of those steps is a seemingly impossible task.

5.2.2.3.2. *The role of religion*

In her discussion of the nexus between Christianity and language learning, Ushioda (2013) states, 'When language learning motivation is intimately bound up with one's Christian faith or sense of Christian vocation, this motivation becomes very powerful and sustaining' (p.224–225). In light of such statements and given not only the Muslim faith of almost all participants, but also their attendance at Islamic schools, the nature of learners' responses concerning religion-related IS images provided noteworthy areas for exploration.

It was assumed that religion would feature strongly in learners' IS concepts, and indeed it was the second most frequently referenced, after travel. The most prominent motivator in the realm of religion was the Quran. This is in line with the findings of other studies, such as Abdelhalim and Alqubayshi (2020), Alkhateeb (2014) and Seymour-Jorn (2004). Learners professed the desire to 'read' the Quran. The subsumption of two potential meanings of this term, namely phonetic decoding and comprehension, under the umbrella of 'reading' at first confuses the picture learners' responses paint. One striking feature of learners' references to religious uses of the language was the use of the term 'when', in contrast to the future-directed, hypothetical 'if' statements used in relation to other IS images. Unlike 'if', 'when' can imply not only a level of certainty of usage, but also a continuity of usage. In the SR Proficiency items, the majority of learners indicated high levels of proficiency in reading words with diacritics (which is how the Quranic script is written). This suggests the majority of learners are already able to phonetically decode this religious text. As such, the desire to read the Quran in the future is less an imagined future state than a continuation of a current one. In fact, when asked about their thoughts on comprehension, many learners reported

their satisfaction with their ability to read, stating that translations were available if they wished to understand what they were reading. This fact may explain why learners did not always reference religion in response to questions explicitly concerning their IS, but would describe religion-related future uses of the language when asked directly about the role of religion in their Arabic learning.

The question of what constitutes a FS image is brought sharply into focus by the integration of the religious component into learners' Ideal Selves. The existence of an 'after-life' component, as well as the overlap between current and desired skills raises the question: Where does a FS begin and end? If an individual wants to understand Quranic Arabic but can already understand some words, and if that individual wants to achieve this skill in order to perform rituals in such a way that gains them reward and earns them subsequent entry into heaven (according to their religious beliefs), is it really possible to delineate a specific FS from which to measure the distance of the CS and determine motivation? Would it, rather, be beneficial to identify multiple FSs: understanding 'x' amount of Quranic Arabic; performing rituals with comprehension of this Arabic; and getting into heaven? Or does this unique dimension of Arabic undermine the transferability and possibly even utility of the L2MSS, by highlighting a major flaw in the distinction between Current and Ideal selves?³⁹

The issue of delineating the future in tighter terms than simply 'any time after now' is an important one. What is noticeable from learners' survey answers and their interview responses are the multiple future reference points in regard to Arabic. When talking about their Arabic-related hopes, fears, and efforts, learners are focused disparately on the GCSE and doing well in the exams, entry into college, ongoing usage of Arabic in understanding the Quran, or a specific goal related to jobs or travel in their adult lives. All of these relate to different, though potentially related goals, not all of which are Arabic-oriented, and which may manifest at different points in the future. These arguably overlap with the 'life domains' and 'speech genres' identified by Nakamura (2019), of which language learned specifically for the sake of an academic credential is a legitimate domain of use. In any case, motivation towards each of these future selves

³⁹ Incidentally, an interesting question is raised as to the nature of the 'self' in the context of an afterlife. Depending on one's theological beliefs, the same 'self' may extend for eternity, or may end when one dies if the self is, at least in part, corporeal.

could, and indeed appears to, differ substantially. Another way around the imposition of both a researcher-generated timeframe and a narrowed IS concept is, as suggested previously, to first ask learners to identify their own Arabic-related IS image, before then answering scale items in relation to it. Important related questions would include when in the future this IS would be realised, and whether or not it is contingent on any other Arabic-related future goals, such as exams or course attendance. Myopic measurements of the future are clearly problematic and reduce what is potentially an eternity to an isolated point of unspecified time 'after now'.

5.2.2.3.3. *Arabic? No, thank you*

Of course, one problem with asking learners to identify their own Arabic-related FS images is that they may not have one at all. A high proportion of participants in this study expressed 'No Arabic' FS, which reflects the 'decidedly lukewarm' attitude to learning Arabic observed by Calafato and Tang (2019, p.33) in their study of adolescent learners of Arabic in the UAE. This finding is, from a pedagogical perspective, concerning. This response was particularly high amongst NHL students and in a number of individual classrooms. The concentration of these responses may be largely explained by the fact that, in some schools, the study of the language is compulsory. Learners, therefore, take the language not by choice but because they are obliged to. This school effect reiterates the necessity of taking into account the role of the learning experience in motivation. The absence of any relevance of Arabic to a future self-image means these learners are not motivated to learn the language, and many spoke of this in tandem with feelings of despondency or frustration. The high frequency of 'No Arabic' responses raises two issues. The first concerns the ethics of obliging learners to take and continue learning a subject towards which they express negative feelings. The second is what precipitated such a stance in the first place. Did negative experiences with learning the subject lead to an absence of an Arabic-specific FS image, or did a pre-existing absence lead to negative feelings, given the consequent irrelevance of classroom content? A pertinent question arising from this is whether negative affective states towards Arabic have the same effect as having no Arabic FS image and the extent to which these states overlap with the terms 'demotivation' (Kikuchi, 2017) and 'amotivation' (Al-Hoorie, 2018). Future research is needed to investigate this.

5.2.2.4. Summary

There are clear distinctions in the strength of IS concept for each of the three language varieties, offering further support for the proposition that they be treated distinctly. Dialect emerges as the most popular, for which HLs and NHLs both have similar strength IS scores. HLs did, however, exhibit higher IS scores for MSA and Classical Arabic at T2, indicating that their IS had strengthened, or had at least been maintained, in these two realms, whereas the same could not be said for their NHL counterparts. If a strong IS is to be considered an essential component of motivation, higher IS scores amongst HLs once again speak to the issue of potential inequality of opportunities in terms of the GCSE for learners without Arabic heritage backgrounds.

Travel, religion, and general skills emerged as the most frequently cited IS concepts, which overlap with the three language varieties. However, the credibility of some learners' answers is belied by their phraseology. The frequent use of hypothetical terms in describing their IS images, particularly in relation to travel and jobs, sheds light on the multiple categories of answers learners give to questions regarding their future selves. The use of 'if' in regard to contingency also highlights the necessity of viewing the future as a spectrum of inter-linked events, rather than identifying a vague future point by which participants will have somehow realised their IS concept. This is not to say learners are not to be believed; rather, it is that their answers reflect layers of nuances which, when analysed, reflect the complexity of the IS.

The question of religion in learners' IS images was also raised. Such references in the interviews were frequent overall but not as oft-cited in response to questions about learners' IS images as anticipated. This is likely due to both the overlap between the present and the future, and the fact that the IS concepts learners expressed related primarily to reading, albeit likely in relation to phonological decoding rather than reading comprehension. This is also supported by the use of 'when' in their answers pertaining to religious usage of Arabic, which indicates a continuity of use rather than an anticipated state they wish to realise in the future.

The above points have implications for research regarding the delineation of the future when it comes to the IS. Learners appear to view their FS as a series of contingent events, of which the GCSE constitutes one. They also indirectly express continuity between their current and future selves in the realms of religion, suggesting the necessity of viewing the future as a temporal spectrum, rather than identifying one timepoint against which learners must compare themselves.

From a pedagogical perspective, the number of learners who expressed no Arabic FS image in the qualitative data is concerning. Regardless of whether or not the IS plays a role in motivation, as described by the L2MSS, the finding that a quarter of learners did not see themselves using Arabic in the future suggests that they see the language itself as irrelevant. It appears that the plethora of uses and benefits of learning a language, both in the short and long-term future, are seldom conveyed adequately to these students. However, in the interests of fairness to the teachers, conveying such images to teenagers does not equate to convincing them of their value.

5.2.3. Ought-to Self

The OTS has been identified by multiple studies as problematic, both in terms of its nature and its role in the L2MSS (see Section 2.1.2.1.2.2). Quantitatively, this construct was measured via the use of four survey items: social pressure, force, disappointment, and consequences, all of which loaded satisfactorily on the underlying latent variable in the CFA. The average-scores for OTS were lower than both the CS and the IS at both timepoints across all language varieties.

In the discussion of the role of the OTS, one potential function is that it constitutes part of the FS, with motivation deriving from the desire to reduce the gap between the CS and the FS. If the OTS is posited to be a component of the FS, the observation that CS scores are higher indicates an inverse gap between the two. This suggests that the OTS, at least in this context, has no such hypothesised motivating power and is potentially redundant.

Nevertheless, the scores for this construct, when compared across varieties, provide an indication of the varying levels of obligation learners find themselves under in relation to different language domains. At both timepoints, contrary to predictions, OTS average-scores for MSA were significantly lower than for the other two varieties. At T1, Classical was also significantly higher than Dialect, but at T2, no significant differences were observed between the two.

5.2.3.1. Dissecting the OTS

The nature of the OTS has been the subject of discussion since the birth of this concept. A number of approaches have been posited, including: (1) a separation of the OTS into *own* and *other* in an attempt to identify the sources of these self-images (see, for example: Higgins, 1987; Lanvers, 2016; Teimouri, 2017), (2) the *anti-ought-to self* (Thompson, 2017) which suggests OTS images are an imposition against which some students rebel, and the *feared self*. This final interpretation was originally proposed as a necessary counterbalance to the IS by Dörnyei (2009) in early workings of the theory and ties in with the notion that learners strive towards an OTS for fear of consequences.

Analysis of the interview data led to the identification of three sources of OTS images: parents, school, and learners themselves. This fits with the idea that this self-concept can be divided into *own* and *other*-sourced images, supported by studies such as Teimouri (2017) and Lanvers (2016). Within each of these categories, however, OTS concepts can be further separated into either those expressed and experienced as: a) a desire to avoid certain consequences, or b) obligations.

Noticeably, the OTS image that is often being referred to is not a FS image involving the use of Arabic; rather, it is the achievement of a good GCSE grade. References to the obligation to become someone who can read modern Arabic texts were non-existent. Some references were made to being able to speak Arabic with family members, though these were usually expressed in the form of an expectation, rather than the explicit avoidance of consequences. There were also references to parental expectations for learners to be able to read and understand the Quran as an important component of the

Islamic faith, though this was rarely couched in terms of obligation or consequence-avoidance, rather as a desire.

Parental influence has been observed in other Arabic-focussed studies (e.g., Alkhateeb, 2014; Donitsa-Schmidt et al., 2004). In the current study, parental expectations took two forms: either for students to take the language at GCSE level, and/or having commenced the course, to do well in the subject. In cases where parental expectations were expressed in terms of performance, it was not always clear whether the emphasis was on learners doing well in the subject in order to obtain a good GCSE grade, or on doing well in the Arabic language itself for the purposes of using it. This was also the case for consequences. Participants spoke of the desire to avoid the disappointment of their parents if they did not succeed in the Arabic GCSE. Again, it was not always made explicit whether the disappointment would be related to getting a low grade (and thus would be the same for any subject) or whether there would be a specific layer of disappointment associated with failure in Arabic. When the focus of these expectations was made clear, the response was predominantly religion-related.

When talking about pressures from school, participants presented these mainly as obligations, since Arabic in a number of schools is a compulsory subject, and even in some schools where it was not, there was no option of dropping the subject once learners had commenced the course. Though the theme of consequence-avoidance was less prominent, some participants spoke of not wishing to disappoint their teachers who had invested time and energy into their teaching efforts.

Whilst both pressure from school and parents to achieve a FS image can be categorised as 'other', participants also spoke of personal feelings of obligations and consequences. The main consequences of a lack of success in Arabic were feelings of personal disappointment or regret, and the effects of poor grades on future plans, such as acceptance into sixth-form colleges. The obligation centred on the necessity of avoiding wasted efforts by transforming the time and energy invested thus far into successful performance in GCSE exams. According to MacIntyre and colleagues (2009), '[t]he sense of losing what one has achieved can be a powerful motivator if the loss is salient to the self' (p.208). This countervailing feared self could potentially provide the 'other half' of

the driving force towards achievement. Whether energy is derived from these thoughts is not clear and would have to be tested, since an opening statement related solely to achievement in the GCSE, to which learners refer almost exclusively in the interviews in the context of the OTS, was not incorporated.

The question that arises from the emergence of these codes—parental, school, and self sources, each divided into consequence-avoidance and obligation—is whether these can be seen to form a cohesive OTS. It may, in fact, be more useful to assume ‘the existence of several, potentially conflicting disparate ought-to L2 self-images in this respect’ (Dörnyei & Al-Hoorie, 2017, p.461). Methods for testing whether this is the case would need to involve robust measures for each potential OTS, which thus far have not been developed, but which present an interesting avenue for research.

5.2.3.2. Religion—The Missing Ought-to Self

Perhaps the most striking finding in the context of Arabic is the noticeable lack of a religious OTS amongst participants. There were only sparse references to learners feeling obliged to learn Arabic for religious reasons. As outlined in the IS section (5.2.2.3.2), religion featured prominently in many learners’ self-concepts, meaning that this component is not absent from the Arabic motivation paradigm. Given the intrinsic nature of Arabic in relation to the religion of Islam, and the fact that the vast majority of participants were Muslim, it was hypothesised that learners would make reference to religion when discussing sources of obligation to learn the language. The absence of this factor in the realm of OTS is, thus, unexpected and discussed further here.

Returning to the distinction between ‘reading’ and ‘comprehension’ of religious texts provides a probable explanation for this finding. The ability to phonetically decode the script allows one to perform daily rituals and engage in other sacred practices (Rosowsky, 2006). The ability to understand the meaning of the original Arabic texts at the individual level could be seen as surplus to these minimum requirements, since comprehension is facilitated (to a degree) by the availability of translations and access to scholarship in multiple languages in print and online. The ability to understand Arabic in order to understand the Quran is not only facilitated by translations but is not

couched in terms of obligation in the interview data. It appears, therefore, that comprehension is perceived less as an obligation than an ideal. Given the religious backgrounds of learners and their attendance at Islamic schools, plus survey results which indicate high levels of self-rated proficiency in phonetic decoding, it is apparent that the vast majority of students have already reached a level of Arabic 'reading' which meets religious requirements. Further, those who speak of comprehension do so in the context of desire, often represented by the use of the phrase 'want to', not expectation or duty, hence the absence of 'have to' in this context. In a population in which phonetic decoding skills have not yet been acquired, or in a population with a broader delineation of what is required religiously in terms of language comprehension, religion may still explicitly feature in learners' OTS concepts. This would need to be explored through further research.

Closer exploration of the data suggests that the 'ought-to' element in terms of religion is not experienced directly by participants, hence the lack of explicit references, but is mediated through the avenues of their parents and the school. As aforementioned, learners make frequent mention to the imposition of Arabic as a subject in their schools (where the GCSE is compulsory), and some mention that their parents sent them to Islamic Schools in order to learn about religion (of which Arabic is considered a component). Whilst this is not mentioned explicitly, the expectation of both the school and parents that their children *should* learn Arabic is implicit in their choice of schools, as well as the schools' choice to make the subject compulsory. When asked why they felt their parents wanted them to learn Arabic, participants referred to the links between Arabic and Islam and the benefits of knowing the language for religious purposes. Similarly, when asked why their respective schools imposed Arabic as a component of the curriculum, students referenced this religious link. It appears, therefore, at least in the context of GCSE Arabic, that sources of the OTS are tangible people and/or institutions, rather than abstract concepts or phenomena, such as religion. For these teenagers, it appears the 'ought-to' needs to be embodied in a physical manifestation of authority with whom they have direct contact, if external, or the threat of an undesired personal self-image which they wish to avoid.

In light of the actual content of the GCSE, as well as learners' experiences of the subject,

an interesting paradox is highlighted: Parents send their children to Islamic schools presumably because of the provision of Islamic Studies subjects. This involves the expectation that Arabic is offered. The school is fully aware of this expectation and seeks to fulfil this by providing Arabic lessons. Learners also clearly express a desire to learn Arabic for religious purposes, amongst others. However, there is no standardised examination assessing religiously-oriented Arabic. Schools are reluctant to dedicate precious resources (e.g., time, staff salaries, etc.) to the provision of a subject without measurable, standardised results, and parents are keen for their children to obtain as many GCSE grades as possible. Thus, the Arabic GCSE curriculum is taught, despite the fact that this does not contain reference to the religious dimensions of Arabic, and nor does it seem that teachers are able to dedicate much time to applying these skills to religious texts during the lessons. As a result, learners are expected to study for an exam which supposedly caters to these original expectations, yet these expectations are, in fact, not being met, and all the while learners are being made to feel stressed and despondent in the process.

Thus, there appears to be an incongruence between the presentation and reality of the Arabic course being taught. Whilst little reference is made to any sense of religious obligation to study Arabic, the sense of religious obligation, or at least expectation, is woven into the schools' choice to offer the language, and the parents' choice of these schools for their children. At the same time, the decision to make it compulsory in many schools, regardless of this initial reasoning, is undermined to a large extent by the limitations of the Arabic GCSE.

5.2.3.3. When is the future, again?

Given the identification of sources of OTS amongst participants, two questions remain, the answers to which are interlinked: why were OTS scores so low, particularly in comparison to the frequency of OTS-related comments in the interviews; and what role does the OTS play in the L2MSS?

One possible reason for the low OTS scores is proposed by Csizér and Lukács (2010) who argue that the absence of significant findings in relation to the OTS was due to the

age of their participants, stating that 'secondary school participants are still relatively young to internalise the pressure the environment might put on them' (p. 6). While this argument appears to have some validity in explaining the low survey responses, it is undermined by the high frequency of comments concerning pressure made by students in the interviews. Comments made with regards to the OTS in the interviews centred around success in the GCSE as their FS reference point. Very few participants spoke of any sense of obligation to pursue any of the FS images as defined by the opening statements of the survey items. The difference in time timeframe of the referents (i.e., the GCSE exams v. the future states described in the opening statements) may explain the disparity. OTS references, it appears, are more prominent in response to shorter term FS concepts. This links back to the question of specifying when the future is. Is it today, tomorrow, after the GCSE exams, in five years' time, a decade, or even in the afterlife? And what is the FS image that is being referenced in that timepoint? In the context of the GCSE, there appears to be a conflation of course and content; when participants are asked about 'doing well' in Arabic, this can mean one of three things: gaining proficiency in the language; gaining a good GCSE result specifically because they want to or should do well in Arabic; and gaining a good GCSE result irrespective of the subject. In educational settings, careful attention must be paid to unpicking these strands since each will entail different IS and OTS concepts, both in the context of Arabic, as well as other school subjects.

The low OTS scores for all varieties were surprising, given both the religious ties to the language, the existence of heritage learners, and the requirements of the GCSE. The finding that the OTS scores were significantly lower than the CS ones undermines the potential motivating power of the OTS as part of the FS, at least in this educational context. This is due to the fact that it does not meet the criteria for any discrepancy between the two selves, which could cause discomfort within students and in turn propel them along a course of action towards realising the selves expressed in the opening statements. The frequency of references to pressure and other OTS-related themes in the interviews, despite the low scores captured by the survey scales, suggests a disjuncture in the FS references being tapped into. Interviewees appeared to be referring to the shorter-term goal of success in the GCSE, whereas the survey scales employed longer-term reference points. This, again, raises the question of delineating

the future. In the shorter term, with success in the GCSE as the FS goal in question, the combined elements of consequence-avoidance and obligation may have some motivational power. This is supported, to a degree, by Higgin's (1998) description of OTS goals as 'minimal goals'. These are usually imposed by others and require minimal fulfilment in order to satisfy the expectations of others. As such, they are unlikely to sustain motivation in the long-run (Al-Hoorie, 2018), which is perhaps why OTS references were made in response to the short-term GCSE goal, but OTS scores were low in response to the longer-term opening statements. Further investigation is needed to determine whether the timeframe of a FS does indeed interact with this self-concept's effects on motivational outcomes.

5.2.3.4. Summary

The findings from the survey and interview data paint a complex, often contradictory picture of the OTS. Whilst it emerges as a distinct construct, it does not appear to fulfil the motivating function it is theorised to have since survey scores were lower than the CS scores across varieties. On the other hand, learners frequently reference pressure and consequences in the interview data, particularly when it comes to studying for the GCSE. This suggests that while the OTS may not play a role in motivation when it comes to the realisation of long-term future selves, but it may serve a function in the shorter term. This distinction in the time-framing of FS concepts offers a further explanation—in addition to the role of language and learning context—for the emergence of often contradictory findings in other language learning contexts (Taguchi et al., 2009). Further investigation is needed to confirm this new hypothesis.

This study goes beyond existing research by incorporating both longitudinal and qualitative components into the research design. The interview analysis provides an in-depth, but also extensive exploration of the components of the construct. Previous qualitative research of Arabic learners' motivations has largely been limited to college or university students (e.g., Anya, 2011; Thompson, 2017; see Appendix A.2 for details), and no studies have employed interviews in the examination of the L2MSS in the context of secondary school learners of Arabic. The incorporation of the interview component into the longitudinal design facilitated the comparison of learners' own self-

expressed FS concepts and those provided in the surveys, which shed light on the importance of factoring in time as a spectrum, rather than only as static points.

A further novel contribution of this study is the analysis of self-concepts in relation to multiple language varieties. In the context of Arabic raises interesting themes regarding the role of religion in the motivational dynamics related to this language. The absence of religion in references to the OTS was unexpected. It appears that either: learners already feel they have fulfilled their religious obligations through learning to phonetically decode the Quranic script; and/or that learners do not experience direct pressure from religion to learn Arabic. Rather, this is manifested indirectly through the expectations of parents and the incorporation of the language into the Islamic curriculum by many of the schools involved.

In terms of the anatomy of the OTS, this was sourced from parents, teachers, and learners' own selves, which supports the distinction made by Higgins (1987) and others (e.g., Blair & Azaz, 2019; Teimouri, 2017) between 'own' and 'other' OTS concepts. This study goes beyond previous research in further dissecting each of these OTS components. Both 'own'- and 'other'-sourced OTS concepts encompass a fear of consequences, which drives effort away from a potential future outcome, and a degree of internalisation of an external expectation or obligation, often manifested simply by the fact that students are taking the subject.

The theme of fear of consequences overlaps with that of contingency. Of note was the fact that consequences of failure (or success) were not confined to the impact on learners' Arabic-specific FS images. A GCSE grade, regardless of the subject, has the perceived potential to affect long-term future plans. Just as the attainment of a FS image is contingent on certain requirements being met, so too can such an image be obliterated by feared selves being realised, rather than desired ones. Conceptually, the trajectory of thought described is illustrated in Figure 5.1:

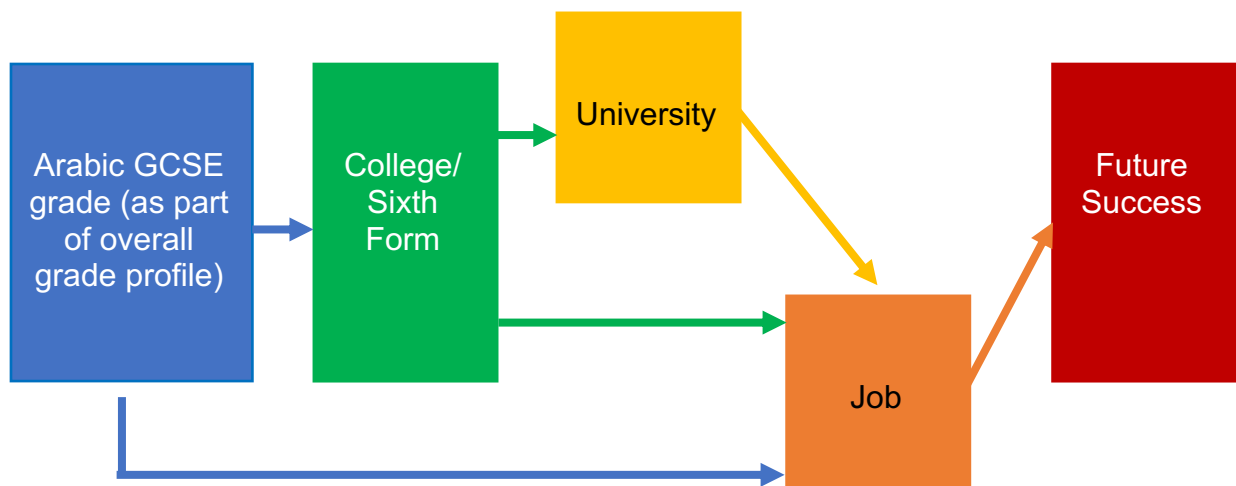


Figure 5.1. Future Self Trajectory

At each stage, there are IS images, expectations, and feared selves— and the theme of contingency runs through each of these. The next step in this ‘flowchart of the future’ is dependent on a certain self being realised and a certain self being avoided. These self-images can be self-sourced, externally imposed, or both. It appears from the interview data that the OTS elements may have some motivational potency within the short term; however, they feature little in the conceptualisation of long-term FS images. The issue with capturing the role of the OTS may stem partly from the focus in L2MSS research to date on two static timepoints: the present, measured once and in-the-moment, and the future, an unspecified forthcoming point at which a narrow ideal or ought-to self is realised. To truly capture the dynamic roles that fear, pressure, desire, hope, and discomfort play in propelling students towards language learning outcomes, a far more fine-grained approach to time needs to be incorporated. This differs from suggestions of incorporating longitudinal measures into research designs, which capture fluctuations in the strengths of self-concepts over time. This, rather, proposes the identification of multiple future timepoints at each stage of the data collection. This will aid in further analysing and understanding the trajectories of FS images that learners may, or may not, map out for themselves and what drives them towards each step in this multi-phased journey.

5.2.4. Past Self

Expanding the scope of self-trajectories should not just be future focused. The L2MSS conceptualises motivation as a force emanating from the desire to reduce the discrepancy between *current* and *future* self-concepts. Restricting the lens on self-concepts to only the present and future has excluded analysis of a vital constituent in the formation of such self-concepts: the *past*.

5.2.4.1. Past Experiences

The findings of this study suggest there are two categories of past self (PS): *past learning experiences* (PLE), and *past personal experiences* (PPE). These PS concepts relate to the CS and FS in different ways.

Past learning experiences relates to participants' experiences studying the language to date, whether in school or as an extra-curricular course, such as the 'Alimah course in local mosques. When operationalised as years of study, PLE was only weakly correlated with five of the nine self-concepts. A function of past learning experiences did emerge in the interview data, through which these experiences can be seen to fulfil the role of comparison. Learners derive their assessment of their CS based on how they feel they are performing in comparison to how they performed in the past, and how confident they are with the material in comparison to their confidence levels with respect to prior course content. Particularly if the CS is conceptualised to include an affective dimension, incorporating measures of the PS is vital to providing context for participants' evaluations of themselves and the content and nature of their study. Despondency, or a lack of confidence, does not necessarily reflect poor current performance, objectively speaking, but may reflect instead higher levels of difficulty experienced with current course contents in comparison to earlier ones.

CS assessments may also be informed by *past personal experiences*, defined as using Arabic in authentic settings outside of the classroom, particularly, such as with Arabic-speaking families or abroad. Evidence for the link between past personal experiences and current feelings towards Arabic study was ascertained via an examination of the

frequency of references to each feeling coded for those with and without past personal experiences using Arabic. Those with no such past experience were four times as likely to report apathy and despondency towards the language, whereas the percentage of those with personal experiences who reported enjoyment of the subject was almost double those without.

By incorporating past experiences, the CS is presented less as a static snapshot, isolated in time and within person, and more as the result of a confluence of factors in relation to which comparisons are constantly being made. Whilst measuring a CS quantitatively at a specific timepoint will force it into a unitary dimension, measuring the PS alongside it adds another aspect to the concept. It also anchors interpretations of the CS in a timeline rather than presenting the construct as a state having sprung up without a source. Stretching the measurement timeline to incorporate past experiences in relation to which the CS is analysed represents a step towards integrating a CDST approach to the study of self-concepts.

5.2.4.2. The Link between the Past and the Future

One prominent, unexpected finding in this study was the link between the past and ideal self. HL participants' existing access to Arabic-speaking communities can be seen as a driver for their IS concepts, given that they are 'already members of such a community' (Olsen, 2018, p. 294), since it provides them with an immediate goal and setting for the application of their skills. However, not all those in the HL category have current access; many of them have had experience of Arabic in authentic settings in the past and still enjoy higher levels of IS. It is, therefore, deemed plausible that NHLs with past experiences may also experience different IS strengths and concepts.⁴⁰ When IS code frequencies were examined based on a segregation of the data into interviewees with and without past personal experiences with Arabic, a number of noteworthy observations were made. Firstly, those with no PPE were over five times as likely to give no IS concept in the interviews, and secondly, learners in this category were more likely to use 'if' phrases in relation to their IS concepts. This was in contrast to those with PPE

⁴⁰ Whilst there was some degree of overlap between the HL and Past Experience categories, there were 30 students who were not classified as HLs who spoke of having had past personal experience with Arabic, suggesting the categories of HL and Past Experience are not synonymous.

who were more likely to use 'when' statements. These findings, though limited in terms of sample size, do suggest a potential link between the existence of past experiences with Arabic and the formation of a genuine IS image. These findings were corroborated, to a degree, by quantitative analyses of the self-concept scores. Mean comparisons revealed that those with PPE had statistically stronger CS concepts across all varieties at T1. At T2, these participants also had statistically stronger MSA IS concepts and higher scores on the Y/N proficiency test.

Theoretically, this link has intuitive appeal. FS images are not constructed out of thin air. Markus and Nurius (1986) themselves postulated that self-guides are derived from past experiences. Logically, the material for a FS image must be supplied before it can be internalised and elaborated. Whilst this may come from hearing about potential future uses for Arabic, it may also come from direct experience with Arabic in the real world. HLs exhibited stronger IS average-scores than NHLs at both timepoints. However, the reasons for this have not been explored in depth. One likely explanation for this difference is that HLs are able to visualise themselves more clearly than their NHL counterparts using Arabic in the future because the context in which they can do so is already being, or has been, experienced. To a lesser degree, it seems, this also applies to some NHLs. Several students mentioned that they had been on holiday to Arabic-speaking countries in the past. Some of these students recounted their experiences trying to speak Arabic during these episodes. An examination of the Arabic FS images of these learners suggested that they enjoyed a more established FS image. This is logical since they have been immersed, even if only briefly, in a context in which Arabic was actually used. This serves to move their FS out of the realm of the purely hypothetical and into the realm of the plausible. A similar situation is apparent in learners' references to the religious application of Arabic. This is clearly a well-established component of their past lives and constitutes an integral part of their FS images, even when they do not explicitly mention it in response to Arabic FS questions. This both corroborates and expands upon Blair and Azaz's (2019) recent research in which they reported direct effects between past experiences of L2 learning in the context of Arabic and intended effort and IS images. The researchers also reported a connection between the valence of past experiences (positive and negative) and the outcome variables, suggesting the necessity of further weaving together cognition (the existence and type

of past selves) and the affective associations of these in analysing their impact on the creation and maintenance of FS images.

The incorporation of the Past Self into the L2MSS, thus, appears to be essential. It contributes both to the construction of the CS, by way of providing a frame of reference and acting as a point of comparison, and to the IS by providing material with which it can be formed and embellished. Pedagogically, the link between the Past Self and IS highlights the acute need for the provision of positive, real-life Arabic language experiences during students' Arabic learning. These have the potential to equip learners with the material needed to construct plausible, vivid FS images that go beyond the generic pictures that learners seem inclined to otherwise present.

5.2.5. Summary

The data in response to the second research question of this study—concerning the nature of learners' motivational self-concepts—were extensive and robust, going beyond previous research in several ways. Findings offer substantial evidence for the need to re-examine the L2MSS in terms of the number of its components and the level of complexity that is incorporated and captured for each.

The CS appears to be a valid and essential addition to the model. Studies to date have typically not incorporated measures of the CS along with both the IS and the OTS, as was done in this research. In addition to measures of confidence and ability, this study highlighted the necessity of including an affective dimension in the measurement of the CS. Methods are also needed via which to contextualise learners' self-assessments in terms of their learning trajectories to date, their views of the language (as distinct from the course), and other subjects they are studying. Measuring the CS at multiple points is essential, but so, too, is capturing its complexity; learners often express nuanced views incorporating seemingly contradictory elements which require attention and analysis. Pedagogically, the fact that HLs maintain higher CS scores throughout the year suggests they enjoy a distinct advantage over NHLs in the GCSE which is not ameliorated over the first year of the GCSE course.

Though responses for the IS corroborate the findings of previous studies in terms of the popularity of travel as a FS image (e.g., Bouteldjoune, 2012; Graham et al., 2016), the nature of IS responses in this study are called into question through learners' phraseology. The word 'if' recurs in participants' answers raising the themes of reliability and contingency, and a conceptualisation of the future as a spectrum rather than a static point in time. In the context of this study, at both T1 and T2, MSA IS concepts were the weakest. Since the GCSE focuses solely on this language variety, this is concerning and calls into question the relevance of the course content, or at least how it is being presented to students.

This research sheds new light on the constituents of the OTS, though its role in motivation remains ambiguous. OTS scores emerged as lower than the CS scores on the survey scales. If motivation stems from the discrepancy between current and future selves then the OTS appears redundant as a motivating force. In addition, learners' responses to interview questions, when compared against their survey responses, highlighted the fact that they may have been referring to different FS concepts: shorter term (GCSE success) in the interviews, and longer term in their survey responses. It may be that the OTS has motivating force in the short term, evidenced the frequency of interview references to it, whereas in the longer term this function diminishes.

Finally, the Past Self is revealed as an integral factor in informing learners' current and future selves. This involved both past experiences with learning, as well as with language use outside the classroom. The findings suggest that incorporating measures of the past self into the L2MSS may provide it with more explanatory power.

5.3. Longitudinal Developments in Self-Concepts and Proficiency

RQ 3: How do learners' motivational self-concepts and proficiency scores develop over time?

In order to address the oversight of much L2MSS-focused research in terms of incorporating a longitudinal view of motivation (Al-Hoorie, 2018), data was collected at two timepoints—at the beginning and end of the academic year. Scores for self-concepts

and subjective and objective proficiency measures at T1 and T2 were compared for the entire sample, and for the subgroups based on gender and heritage language background. Each of these three construct measurements is discussed in turn.

5.3.1. Self-concepts

Comparisons of T1 and T2 average-scores for each self-concept revealed that there were no significant changes over the year for the sample as a whole. This stands in contrast to the findings of Thompson (2007) whose qualitative study captured the 'dynamicity' and fluctuations of her participants' self-concepts. The exception to this lack of change was the Classical CS, which fell, indicating that learners felt less capable in their skills of accessing religious texts. Given that religion was the second most frequently cited reason for learning Arabic, this finding suggests that the learning process has had a negative impact on learners' perceptions of their skills in this important domain. Since their IS concepts did not change, this implies a growing, rather than shrinking gap between learners' current and future self-concepts in this domain. This translates, theoretically, into lower levels of motivation according to the L2MSS. Though this paints a concerning picture, it is perhaps not altogether surprising. Previous studies, such as Graham (2006) and Taylor & Marsden (2014) have also reported low levels of L2 motivation amongst British secondary school students for learning other, more 'traditional' Modern Languages, such as French and Spanish.

In the context of the current study, it is possible that being introduced to a new curriculum at in Year 10 could initially weaken learners' confidence in their own abilities given the shift in difficulty level between Key Stages (KS) 3 and 4, of which learners made frequent mention. However, it is surprising to see that in no way did studying the GCSE course for a year increase learners' IS concepts in any language variety. In other words, as a sample, participants' perceptions of themselves using Arabic in the present and future were not strengthened. The reverse also stands, however; a year of GCSE study did not weaken their Ideal Selves significantly either. Indeed, it is possible to see these results as both a silver lining and the cloud.

Interestingly, and to reiterate the point regarding the distinction between the GCSE as a

FS and the FSs described in the opening statements, OTS scores did not change either. This is despite the frequent mention of pressures and obligations learners reported in the interviews stemming from the need to perform well in the GCSE. Their sense of 'ought to', or obligation, appears to be predominantly tied to grade achievement, rather than the realisation of the types of FS images described in the surveys.

When analysed separately, neither males nor females exhibited different trajectories in the development of their self-concepts over time. This contradicts evidence from other studies suggesting that girls are more motivated to learn languages than boys (Carreira, 2011; Henry, 2010).

However, when it came to comparisons between learners of heritage language background, HLs MSA CS scores rose. Contrastingly, there were no similar significant changes in NHLs' MSA CS scores. In fact, NHLs' CS and IS concepts fell for both Classical and Dialect varieties between T1 and T2. The longitudinal changes between HLs and NHLs, or largely absence thereof requires explanation. It could be argued, again, that exposure to a new curriculum from the beginning of Year 10, drastically different in pace and content from KS3 (during which most of these students were studying Arabic from teacher-generated syllabi and materials)⁴¹ could have an initial weakening effect on learners' self-concepts. This is alluded to in discussions on learners' past learning experiences with Arabic in the interviews, explaining the low-to-middling average-scores at T1 (see Section 5.2.4). This, however, would only explain the initially low scores at T1. The lack of any positive significant changes in these scores over the course of the year, except for HLs' MSA current selves, is concerning on two fronts. Firstly, it is concerning that the initial year of the GCSE does very little to motivate learners, in the sense of promoting IS images and/or decreasing the gap between the Current and Ideal Selves. In fact, for many NHLs it appears to do the opposite in the realms of Classical and Dialect skills. Secondly, in the one domain which could have a significant impact on eventual academic performance (MSA CS), HLs are clearly advantaged over their NHL counterparts. The latter notion reinforces the belief that the GCSE is not equally accessible to all learners regardless of background (Gough & Calderbank, 2019).

⁴¹ There is no national KS 3 curriculum available for Arabic teaching at present.

5.3.2. Proficiency Scores

Comparisons of proficiency test scores revealed that there were also no significant changes in these measures between timepoints. This means that, overall, students improved neither in their vocabulary knowledge, as measured by the Y/N tests, nor their grammatical knowledge, which the cloze tests also tapped into. In other words, a year of studying the GCSE Arabic curriculum did not translate into any significant gains in the language as measured on the tests, despite the fact that the contents of both tests were drawn from the GCSE specifications. The pedagogical implications of this are worrisome, especially given that the GCSE course usually spans only two academic years.

When separated into subgroups, HLs outperformed their NHL counterparts in both the Y/N and cloze tests and at both timepoints. Neither group, however, improved between timepoints. In practical terms, this means that HLs began the year with an advantage in both vocabulary and grammar knowledge and maintained this throughout this year. This corroborates research (e.g., Kagan, 2005) on HLs in other language contexts which suggests the existence of such an advantage. These results pose a significant challenge to the validity of the claim that the same levels of success in the GCSE are ‘accessible to all students’ (Pearson, 2017, p.2).

5.3.3. Self-Rated Proficiency Scores

Subjective measures of proficiency for each of the three varieties were collected via the Self-Rated Proficiency scales included in the Motivation surveys. For the sample as a whole, there were no significant changes between timepoints for any of the items. As with the proficiency tests, there were also no longitudinal changes in any of the items for either gender, nor were there any significant differences between males and females at T1 or T2. Neither HL nor NHL scores changed significantly over time. Unexpectedly, there were also no significant differences in the SR Proficiency items between HLs and NHLs at T1. This indicates that although HLs outperformed NHLs in the proficiency tests, their own perception of their skills did not differ from that of their NHL counterparts, even when it came to skills involving dialect. This could possibly be

explained by their heightened awareness, through exposure, of the skills needed to communicate in spoken colloquial Arabic, which tempers their assessments of their proficiency. By the end of the year, however, HLs clearly have an advantage in terms of their self-perceived language competencies in all language domains. Whether the cause of this increase is to be found inside or outside the classroom is unclear and requires further research.

5.3.4. Summary

Self-concept and proficiency scores were, in a word, static. There were neither improvements in self-concepts nor proficiency measures of either type when the sample was examined as a whole. In practical terms, this means that an entire year of GCSE instruction did not improve learners' CS concepts, their proficiency, or their FS images. A closer look at scores separated by heritage background reveals that HLs' MSA CS concepts rose over time, and that they enjoyed and maintained a significant advantage in their proficiency scores. This means that they perceived themselves as improving in MSA, although in reality their proficiency scores did not reflect this. Nevertheless, they outperformed NHLs throughout the year. Success in the GCSE has long been marketed as accessible to all students. These findings undermine this claim.

Why did learners not improve in their self-concept or proficiency scores? There are a number of possibilities. The first, as aforementioned, is the unfamiliarity of the learners with the KS4 course materials and specification; it may take them time to adjust before they start to see a change in their self-concepts. An examination of Year 11 students' profiles and proficiency would help shed light on whether this takes place over the later span of the course. The second is the perceived unrelatedness of the GCSE to the opening statements in the survey scales. Despite progressing through the course content, learners did not envision themselves as more confident or strengthen their FS concepts because they are unaware of exactly how the GCSE will translate into achieving these goals. Pedagogically, this suggests that the link between school learning and future outcomes is not being emphasised—a missed opportunity if it is the case that strong self-concepts are correlated with motivated behaviour. It may also be the case that learners started the year with misplaced perceptions of their abilities, and that as

the year progressed, these were reassessed. According to Ushioda (2014), the levels of motivation with which learners commence courses are often based on expectations—which may be unrealistic—rather than an awareness of what is actually entailed. The readjustment of participants' self-concept scores is as much as a reflection of changes in their perceptions of the course requirements as it is of their self-perceptions. In other words, relative to what they understood of the course at the beginning of the year, they may be more confident, but relative to what they now know, they are less confident. The adjustment in expectations may also reflect learners' waning 'optimism that [future] competences can be developed' (Spinath & Steinmayr, 2008, p.1567). The final possible explanation pertains to the actual progress they made over the year—it may be that learners did not cover enough material by the time of the second survey for it to have made any impact on their self-concepts. If this explanation is accurate, however, this would have concerning implications. The second set of surveys were completed at the end of the first of the two academic years, Years 10 and 11, during which GCSE content is covered. A lack of significant progress during a whole half of the course means that either learners are likely to receive generally low grades in their GCSE exams (NHL learners in particular, since their proficiency scores were significantly lower than those of HLs) or that intensive effort is required during Year 11 in order for learners to improve. It is possible that adjustments to teaching approach or methods of instruction could address this lack of progress. Future studies incorporating measures of teaching practice are therefore recommended.

5.4. The Relationship between Proficiency and Self-Concepts

RQ 4: What is the relationship between proficiency and self-concepts

a) Within timepoints?

b) Between timepoints?

In order to investigate the relationships between self-concepts and the two measures of proficiency, path model analyses using factor scores generated from the CFA were run. The literature to date offers evidence for the predictive power of the IS in regard to outcome measures, most commonly operationalised as Effort, or Intended Effort (see Section 2.1.2.1.2.4 for discussion on criterion measures). However, in this study, no such

relationship between the IS and outcome measures emerged for any of the three language varieties. This corroborates preliminary findings by Al Khalil (2011) who found no association between L2MSS measures and oral production. A new 'proficiency-driven model' was observed in the current study, in which proficiency test scores at T1 predicted changes in self-concepts at T2. The conceptualisation that earlier skills predict positive change in self-concept is consistent, to a degree, with the 'skill development model' of Marsh and colleagues (2016) observed amongst school-aged learners. The new measurements of proficiency and the proficiency-driven model, both novel contributions to the field, are discussed below.

5.4.1. New Outcome Measures

It is important to note in discussing the findings of this research question, that this study incorporated two operationalisations of outcome measures: a set of subjective self-rated proficiency scores, and objective proficiency tests measuring learners' lexical and grammatical knowledge. Rather than measuring intended motivational behaviour, these represented measures of the results of motivated behaviour, i.e., the outcomes of effort expended. These alternative outcome variables may offer an explanation as to why very different patterns emerged in this study in comparison to previous research (see: Al-Hoorie, 2018). The path model analyses indicate that, although they remain stable, future self-concepts at T1 have little to no predictive power upon the proficiency outcome measures, either subjective or objective. This confirms the findings of studies such as Papi and Abdollahzadeh (2012), in which classroom behaviour was not predicted by the IS, and Moskovsky et al. (2016), who suggest there is 'at best [...] a tenuous link between the self guides and achievement' (p. 650). Rather than self-concepts predicting outcome measures, it appears instead that proficiency scores of both types predict self-concepts over time, and in varying ways dependent on the language variety.

These results have two important, immediate implications: The first is that treating language varieties as separate (though inter-related) entities, in relation to which learners have distinct motivational profiles, has empirical support. These varieties can be combined into an overarching 'Arabic' construct; however, there are important

dynamics which can only be explained by examining the nuances of each variety. The second, though tentative, implication is that a self-concept driven model of motivation, when proficiency measures are the outcome variable, is not supported in this context. What the findings instead illustrate is a proficiency-driven model, whereby language skills (both self-rated and objective) at T1 predict self-concepts at T2 in various ways, illustrated in Figure 5.2.

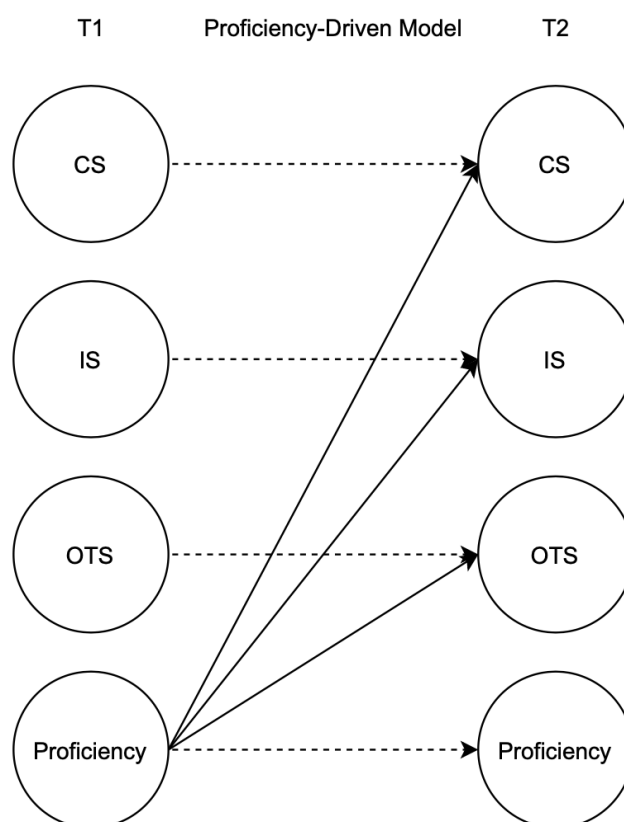


Figure 5.2. The Proficiency-Driven Model

5.4.1.1. The Many Faces of Proficiency

Proficiency in this study is measured bi-dimensionally, consisting of two sets of skills. Close inspection of the components of proficiency sits in line with a CDST approach. Proficiency as a construct is presented as multifarious, and its interactions with learners' self-concepts are clearly complex. While this study certainly only goes so far in capturing its breadth and depth, it represents a step in the direction towards a complex dynamic systems approach by seeking to incorporate the subjective and the objective, each as a multifaceted construct.

The treatment of proficiency as multi-componential is supported by the emergence of different interactions between each one in relation to the hypothesised predictor variables. In terms of the proficiency test measures themselves, it appears that grammatical knowledge (measured, in part, by the cloze test) contributes both to the development of subsequent grammar skills and to building a stronger lexical repertoire. The reverse is not true; vocabulary knowledge, measured by the Y/N test, does not significantly predict changes in cloze test scores. This may be because Arabic grammar involves an extensive derivational morphological system, whereby trilateral roots can be transformed, according to specific and very regular templates, into an expansive host of nouns and verbs. Thus, learning morphological grammar would enhance learners' abilities to identify, at the very least, the roots of words and extrapolate meaning from these using their knowledge of derivational systems (Abadzi & Martelli, 2014).

Papi and colleagues (2018) reported the OTS/own category to be to be the strongest predictor of motivated behaviour. Whilst there is precedence for the link between outcome variables, an unexpected finding in this study was the inverse relationship between the OTS and proficiency. A negative correlation coefficient (-0.5) between the OTS and achievement was also found by Al-Hoorie (2018) in his meta-analysis of L2MSS studies, though no explanation is offered for this particular result. There are, thus, two questions here: a) Why is it that a higher proficiency at T1 would result in a weaker OTS concept at T2, and b) why is it specifically between the grammar-oriented proficiency test and OTS that an inverse relationship emerges? One potential explanation to the first question lies within the hypothesised underlying preventative nature of a learner's OTS. If a person is performing well at T1, it is possible that they feel less pressure to achieve a FS state in order to avoid negative consequences. Their initial higher level of proficiency diminishes the potential of negative consequences from the outset. By way of example, if parents perceive their child to be performing well in the language, they may feel less need to exert pressure upon them to achieve a future goal.

It is possible that grammatical knowledge is a stronger indicator of underlying confidence in one's language skills, to the extent that it diminishes the degree to which learners feel pressure to avoid negative future outcomes. A stronger grasp of

grammatical knowledge in the context of Arabic, given its derivational root system, may promote vocabulary gains and afford one a level of confidence which ameliorates the need to be vigilant in avoiding negative future consequences, and the building of a future self around these threats. Further exploration of this hypothesis is needed. There may also be underlying classroom effects at work; classes in which learners express lower OTS overlap with those with higher levels of grammatical proficiency. Whether there is a causal link between the two requires further investigation.

5.4.2. Linear or Reciprocal?

A degree of precedence for this emergent proficiency-driven model can be found in the work of Marsh and colleagues in their exploration of the relationship between academic achievement and academic self-concept (see, e.g., Marsh, Byrne & Yeung, 1999). Initially, they posited two models: *self-enhancement* and *skill-development*. In the former, achievement is driven by self-concept. This would be the underlying assumption of the original L2MSS. The latter implies the reverse; that self-concepts, in fact, are driven by academic achievement. The assumed self-enhancement model underlying the L2MSS does not emerge in this study. The skill-development model, as with the proficiency-driven model proposed in the current study, suggests that learners' self-concepts are formed on the basis of their actual achievement and will be affected by developments in their objective test scores. The two models—self-enhancement and skill-development—were later integrated into a reciprocal-effects model whereby the two interactions took place simultaneously (Marsh et al., 2016), aligning closely with a CDST approach to motivational systems.

Though Marsh et al.'s (1999) skill-development model provides a useful prism through which to interpret these emergent results, caution must be applied in stretching its application. Firstly, the researchers do not examine self-concepts in the context of language learning, which are arguably more complex than those in relation to mathematics (the focus of their study) given the strong associations between language and identity (Norton, 2020). Secondly, they do not examine future self-concepts, only current academic ones. Further research is required to ascertain whether a skill-development model would emerge in language learning contexts and whether the

relationship between achievement and FS, rather than CS, concepts also exists. To say, therefore, that a skill-development model emerged from this data would be misleading; hence the use of the researcher's own phrase 'proficiency-driven'.

In addition, the substantial reciprocal relationships posited by Marsh and colleagues did not emerge in the path models run in this study. However, a reciprocal effects model cannot be ruled out completely in this context. It may well be the case that subsequent proficiency at a later timepoint is predicted by T2 self-concepts. Wigfield and Karpathian (1991) argue that perceptions of ability need to first be well-established before they assume any predictive power in the realms of achievement. Since most learners in this study had only just begun the GCSE, their self-concepts in this particular domain may not be well-grounded enough at T1 to be motivating. Further studies are needed to determine whether a reciprocal effects model emerges across a longer timeframe.

In sum, the findings of the present study suggest that for learners commencing the Arabic GCSE, IS concepts are not the driving force behind any improvement in their actual behaviour-dependent outcomes, i.e., proficiency. Simply put, the variability in their vocabulary and grammatical knowledge and their SR proficiency is not explained by the strength of their IS concepts earlier in the year. Their CS concept, however, does play a limited role in predicting vocabulary gains over time. Between the CS concept and both types of proficiency measures there seems to be a degree of reciprocity, which makes sense since proficiency—actual and perceived—informs learners' levels of confidence and ability as measured by the CS, and vice versa. Given that the level of proficiency learners exhibit upon entering the classroom appears to be the driving force behind any developments in their self-concepts and actual skills, this makes the question of fairness all the more pertinent in light of the significant differences in proficiency between HLs and NHLs.

What does this mean for teachers wishing to boost learners' motivation and motivation-related outcomes? It appears that how well learners are actually doing, particularly in the realms of vocabulary knowledge, has the strongest impact over time on how likely they see themselves using Arabic in the future. To boost proficiency over time, learners

need to see evidence of themselves as confident, able, and actual users of the language, as measured by the CS, which was the only self-concept predictor of proficiency measures at T2 in any language-variety model. Knocks to learners' CS concepts are, thus, likely to have consequences for achievement. This is possibly why so few gains were made amongst learners in this sample in proficiency skills over time, given the trend of a lack of confidence perpetuated by the perceived difficulty of the new 9-1 specification and, to some degree, the perception of the language itself.

5.4.3. School-Level Differences – The Role of the Learning Experience (L2LE)

The L2LE has been conceptualised as an important contributor to learners' motivations, and previous studies have found some support for this (Al-Hoorie, 2018). Whilst an exploration of the L2LE did not constitute a focus of this research, preliminary investigations into differences at the classroom-level revealed interesting findings that speak to the importance of the learning experience in language learning motivations.

Learners were clustered into 19 classrooms at T1 and 13 at T2.⁴² Intra-Class Correlation (ICC) coefficients indicated high levels of class-based variability in the context of self-concepts and proficiency measures. ICC scores were particularly high for the CS concepts, indicating that classroom factors played a substantial role in explaining a large proportion of the variance in these variables. If affective factors are posited to be a component of the CS, this would make sense. The existence of 'group-level energy', or 'emotional contagion', refers to a process whereby emotions spread through a group and result in a 'collective motivational level' (Al-Hoorie & MacIntyre, 2020, p.15). Since clusters of references to particular emotions were observed in different classrooms, it is possible that such a phenomenon had occurred. ICC scores for the IS were also high. If the IS does play a role in promoting learning, then this is a promising finding for teachers, since it demonstrates the potential for classroom factors to promote the development of learners' IS concepts. At the same time, a cautionary message is implicit: classroom factors can also contribute to low IS scores. The identification of the exact mechanisms behind this promotion/erosion function of classroom factors, and precisely

⁴² 'Classrooms' are referred to as the grouping variable rather than 'schools', since in one school, girls and boys were taught separately by two different teachers.

which classroom features are at work, requires further research.

Contrastively, the ICC scores for OTS were low, indicating that classroom-level factors played a less pronounced role in forming learners' self-concepts in this domain. This suggests that, in the realms of Arabic language learning, classroom factors play a small role in imposing either expectations or consequences when it comes to learners' longer-term FS concepts. This is in contrast to the pressure that can be exerted on students by teachers and schools (e.g., by making subjects compulsory) in the context of shorter-term academic-oriented goals, such as passing the GCSE, as is evidenced in the interview data.

This insight into the existence of classroom-level differences further demonstrates the necessity of factoring in measures of the L2LE in motivation research, largely overlooked to date. Though Dörnyei (2019a) has recently suggested conceptualising the L2LE as engagement, which was indeed the initial approach intended for this research at its outset, it seems from the interview data in classes which exhibit significantly different levels of CS and IS concepts that factors beyond task engagement affect learners' self-concepts and different elements of their proficiency. Controlling for classroom-level differences was not possible in the path model analysis given the sample size and number of free parameters. Future research is needed to determine the extent of inter-classroom variability and the effect of this on the model outcomes.

It requires mention that classrooms overlapped largely with schools, since in the majority of participating schools only one Arabic class was taught. The school-level factor would also incorporate demographic features, such as school type, the socio-economic status of students, and even the predominance of Arabic spoken in local communities. Whilst examination of these factors was beyond the scope of this research, the interplay between such school-level variables, classroom factors (such as teacher and teaching style), and individual learners presents a further, necessary and promising area for future research.

5.4.4. Summary

Though the structural validity of the L2MSS was confirmed by the CFA earlier in the study, the dynamics and explanatory power of this model are challenged by the analyses in response to RQ 4. The model proposed by the literature, in which the IS predicts outcome measures over time, did not emerge. Rather, the reverse was apparent. Proficiency measures predicted self-concepts over time. This model is not a reciprocal effects one, as might be anticipated. It may be that such effects did not emerge given the lack of statistical differences between the means for the factors at T1 and T2, necessitating further studies, extending over longer time periods, to test these findings. What is noteworthy is the fact that different measures of proficiency exhibit different relationships with self-concepts over time, supporting their separate analysis, and the complexity of proficiency as a construct. Furthermore, the effect of classroom factors highlights the role of the L2LE and the impact that classroom-level variables can have on learners' self-concepts and proficiencies. Measures of the L2LE are crucial for future research.

In incorporating an examination of reciprocal relationships between motivational self-concepts and proficiency in the context of Arabic language learning in UK schools, this study contributes to the field both methodologically and in terms of the language learning context. The broad implications of these findings are twofold. In classrooms, emphasis must be placed on improving proficiency in order to promote developments in self-concepts. In research, emphasis must be placed on ascertaining the role of self-concepts in the context of motivation if they do not predict outcome measures.

5.5. The Role of Self-Discrepancy

RQ 5: What role does self-discrepancy play in the L2MSS?

5.5.1. Overview of Findings

On the basis of prior research, one may be mistaken in assuming that the IS is the sole driving force behind learners' motivation. Whilst the IS has, in some contexts, been reported to be a predictor of intended effort (see Section 2.1.2.1.1), the central tenet of

the L2MSS and the theorised source of motivated behaviour stems from the discrepancy between one's CS and IS and the desire to reduce it. The gap between the two, it is thus postulated, determines how motivated learners actually are (Dörnyei & Ryan, 2015).

Following Higgins (1987) who originated the theory, self-discrepancy (S-D) was measured indirectly, and was operationalised as the numerical difference between the CS and the IS. Discrepancy scores were first calculated for each language variety at both timepoints from the average-score values. At T1, the S-D scores for Dialect and MSA were significantly higher than for Classical. This indicates that learners felt closer to achieving their IS concept for the Classical variety. Based on previous discussions concerning the current level of usage of, and exposure to, Classical Arabic through religious practice, this is unsurprising.

At T2, Dialect Discrepancy scores were significantly higher than both MSA and Classical. This is expected given the finding that IS scores for Dialect were markedly higher than for other varieties at both timepoints. Comparisons between the S-D scores over time revealed that the discrepancy increased for Dialect, but not for the other two varieties. It may be the case that learners' perceptions of themselves as users of Dialect had weakened, or that their IS concepts had strengthened in comparison to their current self-perceptions. Further research employing other statistical analyses, such as Latent Profile Analysis, would aid in ascertaining which of these took place.

At T1, NHLs' S-D scores were significantly larger than HLs' for all varieties. At T2, significant differences emerged for MSA and for the combined Overall higher-order construct scores, indicating that NHLs' scores were larger than those of their HL counterparts for only this variety. This is largely explained by the significant increases in HL's MSA CS scores over time. This indicates that over the year, HLs made progress in reducing the gap between their CS and IS scores, whereas NHLs did not. In the context of the GCSE, the fact that HLs exhibit an empirical advantage alludes to the existence of different learning trajectories that need to be taken into account in the delivery of Arabic instruction and the development of course content.

Path models were run to investigate the interactions between the S-D and proficiency

variables. Unexpectedly, S-D was neither predicted by any of the proficiency measures across timepoints, nor did it predict them. This appears to undermine the hypothesised role of S-D in motivating behaviour designed to reduce the gap between current and future self-concepts, which here is measured via proficiency levels.

5.5.2. The Optimal Gap Hypothesis

Could this absence of a linear link between S-D and the outcome measures be explained by the 'optimal gap' hypothesis? This theory is based on the descriptions of the theoretical framework of the L2MSS but has been distilled and labelled by the researcher in this current study. The theory purports that there is a window of optimal distance between the CS and IS, within which this gap is considered motivating, but before and after which the discrepancy is either too small or too big to be reconciled through motivated behaviour (see Section 2.1.2.1.2.3). If conceptualised as such, when plotted against outcome measures, S-D curves should be quadratic, shaped like a bell-curve, rather than linear. When these were plotted, however, no pronounced bell-curve relationship was apparent. In fact, scatter plots exhibited slight inverse quadratic curves, which were unexpected. The plots suggest that higher and lower levels of S-D result in higher levels of proficiency, and that participants with middling S-D scores at T1 went on to have lower levels of proficiency at T2. The R^2 scores for these curves were low however, and suggest that this pattern explains little of the variance in the proficiency scores.

A further attempt to identify the role of Self-Discrepancy in predicting outcome measures saw this variable modelled along with changes in discrepancy to explore whether or not the size of the discrepancy predicted the amount of gains learners made over the academic year. Again, no significant results emerged.

5.5.3. Summary

The response to the question 'What role does self-discrepancy play in the L2MSS?' thus remains unclear. The construct appears to play no linear role in predicting the dependent variables across time. Nor does it appear that there is a quadratic

relationship at work. Considering the supposedly integral role of self-discrepancy in the L2MSS (Dörnyei, 2005), these results are surprising. If these findings were to be replicated in future studies, their implications for the field would be substantial: in the context of Arabic secondary school learners, not only does the IS have little to no predictive power in regard to proficiency outcome measures over time, but the operationalisation and role of self-discrepancy is called into question. Despite the intuitive appeal of the theory, for it alludes to the familiar discomfort of being at odds with the ideal image of ourselves, empirical grounding is lacking, at least in the context of secondary school learners of Arabic. Further ways to explore this are theoretically possible, and present interesting areas for future research.

5.6. Summary: The Future Self of Motivation or the Future of Motivational Selves?

To summarise, the tri-partite structure of the L2MSS model of motivation—reconceptualised to include the CS, IS and OTS—was confirmed by the analyses conducted in RQ 1 in the context of each Arabic variety and when scores were combined into an ‘Overall’ (higher-order construct) model. The different factor loadings for each variety, and the emergence of significant differences between the self-concepts for the individual language varieties (MSA, Classical and Dialect) supports the decision to treat Arabic as a multiglossic language.

However, although initial support for the motivational self-concepts emerged via the CFA, detailed interview analyses and path modelling in response to subsequent research questions reveal challenges to both the conceptualisation of the components—which require a far more fine-grained approach—and the underlying mechanisms of motivation, which are called into question.

There is support for the existence of CS and FS concepts amongst participants in this sample. However, there appears to be a need to expand the scope of these self-concepts further to incorporate past selves and to broaden and deepen the identification and measurement of current and future selves. At the same time, the interactions between these self-concepts and the outcome measures, as well as the role of self-discrepancy, are called into question. Findings suggest that self-concepts are promoted by

proficiency over time, but that these self-concepts themselves do not contribute to developments in proficiency. Similarly, self-discrepancy—the gap between CS and IS concepts—does not appear to play any role in promoting developments in proficiency outcome measures in the context of secondary school learners of Arabic.

From a pedagogical perspective, the picture painted of these learners of GCSE Arabic is arguably worrisome. Though there were positive comments in the interviews about learning the language, learners also report levels of disinterest in Arabic, feelings of despondency in response to the learning process, perceptions of the language (and/or the course) as difficult and appear to be unaware of the transferability of their skills to authentic contexts of use. Furthermore, learners' self-concepts were not strengthened over the course of the academic year, nor was there any improvement in their scores on proficiency tests (which were developed directly from GCSE materials). Heritage learners maintained higher scores in both self-concepts and proficiency skills throughout the year, offering counterevidence to the exam board's claim that the course is equally accessible to all students (Pearson, 2017).

6. CONCLUSION

This chapter details the key findings of this study in the context of both L2 motivation and the learning of Arabic. The strengths and contributions of the research are presented along with a discussion of its limitations. This is followed by an outline of the implications of the findings of this study for pedagogy and practice. The chapter concludes with important areas for future research.

6.1.1. Key Findings

Models are defined by both their component parts and the interactions between them. This study set out to examine the validity and nature of the components of a revised model of the L2MSS—consisting of Current, Ideal, and Ought-to self-concepts—and the relationships between these constructs, and with proficiency outcome measures. This was done in the under-explored context of UK secondary school learners of Arabic, defined as a multiglossic language and with close religious ties.

6.1.1.1. L2MSS

In relation to the revised model of the L2MSS, this study offers key findings in relation to both the strands mentioned above: the nature of the self-concepts; and the interactions between these and the outcome variables. In terms of the first strand, the CS, IS, and OTS all emerged as distinct concepts when measured using the motivation survey scale items at both time points and across each of the three language varieties examined. This confirms the work of other researchers who have also found empirical support for these self-concepts (see Section 2.1.2.1.1). However, findings from the interviews in particular highlighted issues and unanticipated complexities involved in each of these self-concepts, as well as the necessity of including a fourth self-concept component, the Past Self.

One major finding in relation to the first strand is that the CS is far more complex than envisioned. Although the CFA provided support for the construct, the interviews

revealed a more nuanced view of the CS (Section 4.2.6). Findings suggest that affect plays a role in the formation of the CS, a dimension which requires exploration in future studies. Furthermore, this self-concept appears to be formed through comparison with both past learning experiences and other current learning experiences. It is also informed by learners' perceptions of the language and is often multifaceted, incorporating both positive and negative self-appraisals and perceptions.

Learners' IS concepts, it emerged, are often contingent on shorter-term goals and achievements themselves constituting FS images (Section 4.2.7). The future is thus less a specified point in time than a spectrum. Moreover, IS concepts expressed by learners are not always internalised. The degree of internalisation is often alluded to via the use of uncertain, or conditional statements. The IS, it seems, is not always ideal but rather hypothetical, and close analysis of learners' responses to IS-oriented questions is required to determine the extent to which the images described are genuinely part of their future self-concepts.

A number of key findings emerged in the context of the OTS. Evidence was found in support of the division between own- (self) and other- (parents and school) sourced images, corroborating the work of Higgins (1987) who proposed this distinction and Teimouri (2017) who explored this bifurcation in more recent contexts. This study, however, found that the OTS was further divided into themes of consequence avoidance and expectations (Section 4.2.8).

The necessity of conceptualising the future as a spectrum, rather than isolated, static timepoints is also supported by findings in relation to the OTS. Learners spoke of OTS themes in relation to shorter-term FS concepts, predominantly the GCSE, whereas their OTS scores were low in response to the opening statements in the surveys. This suggests that self-concepts vary depending on which point along the future spectrum learners are focused, and that whilst the OTS may have some motivating force in the short term, in the longer-term its role is unclear. This is in addition to the fact that learners' OTS scores were lower than their CS ones in relation to each of the three varieties, further calling into question its role in promoting learning within the framework of the L2MSS.

Perhaps the most prominent finding in relation to the components of the L2MSS is the role of the Past Self. This consists of past learning experiences and past experiences using the language outside the classroom. This confirms and expands upon the findings of Blair and Azaz (2019) who observed a link between FS concepts and episodic memory. Findings from the current study go further, suggesting the Past Self needs to be incorporated into the model of motivation since it contributes to the contextualization of the CS and the availability of the IS (Section 4.2.9). Finally, it was observed that each of the three self-concepts are influenced to differing degrees by classroom factors, adding to calls for measures of the L2LE to be integrated into future studies (Section 4.4.5.2).

Important outcomes were also observed in relation to the second emergent strand—the relationship between motivational self-concepts and developments in proficiency over time. Most notably, support for the predictive power of the IS in relation to outcome measures was not found in this study. In fact, no self-concepts predicted proficiency—self-reported or test-based—over time. This is significant on two counts. Firstly, this contradicts the findings of numerous studies which have found links between the IS and criterion measures (see Al-Hoorie, 2018). Secondly, the absence of any significant predictive function of the IS may be explained by the use of the objective outcome measures in this study. Moskovsky et al. (2016) assert that ‘therein lies the real test for the theory—in the capacity of the self guides to predict L2 achievement’ (p.643). It appears, at least in the context of this research, that the L2MSS fails this test.

Few studies have delved deeper than the self-concept components of the L2MSS in their examination of the workings of the model. To address this, the nature and role of self-discrepancy (S-D) was examined. Although Thorsen et al., (2020) found that this construct did predict intended effort, no such relationship between S-D and the outcome measures was found in this study. Both linear and quadratic relationships were explored, employing alternative outcome measures. The results of these investigations provide no evidence for the role of self-discrepancy in promoting proficiency outcomes. These findings challenge the underpinnings of the L2MSS and call into question the mechanisms underlying the theoretical model.

What, then, are the dynamics at work? Self-concepts were predicted by their own counterparts over time, demonstrating longitudinal stability in these constructs. Most importantly, proficiency test variables at T1 were observed to predict changes in self-concepts at T2, though reciprocal effects did not emerge. This is labelled the *proficiency-driven model*. Unexpectedly, different measures of objective proficiency measures in this model predict different self-concepts. Learners' IS and CS scores at T2 are predicted by vocabulary knowledge at T1, whereas the OTS regresses negatively onto cloze test scores, indicating that the higher a learner's broader (lexical and grammatical) proficiency at T1, the weaker their OTS at T2.

6.1.1.2. Arabic Language Learning

The current study also makes a number of substantial contributions to the field of Arabic language learning, and the context of the Arabic GCSE in particular. Firstly, different factor loadings were observed for each of the self-concepts for each Arabic language variety, and there were significant differences between the average-scores for each. This provides validation for the conceptualisation of Arabic as a multiglossic language, rather than a monolithic entity. Both quantitative and qualitative findings support the conceptualisation of Arabic as multiglossic (Section 5.1.2.2), and learners are observed to interact differently with each language variety in terms of their motivational self-concepts. They appear to be most cognizant of the differences between MSA and Classical in application; however, they exhibit very little awareness of the utility of MSA in real-life contexts and the necessity of Dialect for communicating in everyday spoken contexts.

As pertains to interactions with these varieties, learners clearly conveyed different CS and IS concepts in relation to each. This further substantiates the need to conceptualise these varieties separately. Learners' MSA-related self-concepts are weakest, despite (or because of) the fact that they are learning this variety in their classrooms for the purposes of the GCSE (Section 5.2.2.1). Learners exhibit stronger IS images for Dialect and Classical Arabic, calling into question the relevance of the course content.

Two crucial findings in the context of the GCSE in particular pertain to learners' progress and the issue of course accessibility. Results revealed that learners in this sample did not improve in either their self-concept or proficiency scores over the course of an entire academic year (Section 4.4.1). Furthermore, heritage language learners develop and/or maintain a significant advantage in both self-concepts and proficiency over the year, calling into question claims that the course is equally accessible to all learners. Against the fact that many of these learners are obliged to take the subject by their schools, these results are concerning.

The Arabic language is closely tied to the Muslim faith. Religion was thus hypothesised to play a role in learners' self-concepts. Contrary to these expectations, religion does not feature directly in learners' OTS concepts. Rather, it is manifested indirectly through the expectations of authority figures. On the other hand, religion does play a role in learners' IS concepts, constituting the second most frequently cited category after travel in learners' self-expressed Arabic FS concepts. The exploration of religion in relation to motivational self-concepts revealed the overlap between learners' CS and IS concepts, and the expression of tempered self-concepts, limited to the ability to phonetically decode the Arabic script rather than to fully comprehend the language (Section 5.2.2.3.2).

6.1.1.3. Re-envisioning the Motivational Self System

The question now stands: What is the future of the L2MSS? Though the findings of this study certainly require corroboration, as they stand they present serious challenges to the model. There are two possible routes available: retire the L2MSS in favour of another model of motivation; or reconceptualise the model in the light of these findings.

The latter approach would require an incorporation of the essential element of time, and methods for capturing the 'web of relationships that continually grow, change and adapt to new situations' associated with a CDST approach (Hiver & Larsen-Freeman, 2020, p.289) in order to fully grasp the components and interactions at work. Any such revised model would need to capture the future as a spectrum along which multiple, staggered self-concepts exist, and which are each contingent on those before them.

These FS images may be ideal or ought-to, though it is not clear how far the role of the latter extends. The Past Self, also needs to feature, in particular experiences with the language in context, since this contributes to the availability and plausibility of FS images in either their respective, or indeed multiple domains. Past experiences with Arabic in each domain also inform learners' levels of confidence and their ability in the present, which also informs the construction and plausibility of learners' FS images. The greater the overlap between temporal self-concepts, the more vivid and plausible the FS concept is expected to be. As such, this constitutes an ontogenetic model of second language learning motivation, which incorporates 'accumulated accounts of personal, microgenetic "aha" moments, which require substantial longitudinal analysis of individual language learners as unique agentic beings' (Al-Hoorie, 2021). This results in the creation of a personal timeline for subjects to allow for the in-depth analysis of motivational trajectories.

According to Hiver and Larsen-Freeman, 'Contextual factors are actual dimensions of the system itself' (2020, p.294). Though direct measures of the L2LE were restricted in this study, the incorporation of these elements into the theory is of paramount importance going forward. A 'person-in-context' approach has long been proposed by Ushioda (2009). The findings of this study expand upon the incorporation of the interaction between person and context, by examining the 'person-across-context' (which Donato and Davin (2018) refer to as 'history-in-person')—in other words, an ontogenetic examination of a learner throughout their personal timeline, from their past to their projected selves. The learning experience is not time-constrained; past learning experiences must also be captured, in addition to hopes and aspirations, so as to ascertain their valence and determine the nature of the internalised context which learners bring with them to the classroom. While such a conceptualisation may seem overly complex, oversimplifications of motivation—which attempt to reduce the driving force behind achievement to the strength of an IS or the gap between the CS-IS—failed to capture the complex systems at work within the learner and the learning context in this study. The mechanisms at work are clearly highly nuanced and context-based, and require a fine grained approach in order to identify precisely what is at work when a learner is motivated and when they are not.

6.1.2. Strengths of the Study

In additions to the findings of this study, which constitute a number of novel contributions to the field, several of the strengths of this research also lie within its design. Firstly, in contrast to the majority of Arabic motivation studies, and many L2MSS studies in general, the instruments developed for the purposes of this study were not only grounded in an extensive review of the literature, but also validated empirically (Section 4.1). The Motivation Surveys in particular can be adapted and used in multiple contexts to both examine the generalisability of these findings and further refine their design. In addition, the longitudinal design of the research offers a much-needed perspective in the field of L2MSS studies, which is dominated by cross-sectional designs. This facilitated an examination of motivational self-concepts as dynamic constructs. The incorporation of multiple timepoints also allowed for an investigation of reciprocal-effects between motivation and outcome measures, which is lacking in the field. Perhaps the greatest strength of this study was its mixed-methods approach. While this is a growing trend, most studies superficially combine quantitative and qualitative components (Boo et al., 2015), omitting the integration of datasets that was pursued in this work. Findings from the close examination of interview and survey data were enhanced by the interweaving of these datasets by following threads from one into the other in the pursuit of answers and explanations to arising questions. This is consistent with a CDST approach, which places emphasis on capturing the complexities of systems at work through the application of multiple lenses (Section 2.1.3).

Two further strengths of this study concern its focus. It addresses the dearth of studies on languages other than English (LOTes) in the literature (Section 2.1.2.1.2.5) and provides a unique perspective into motivation by focussing not only on a multiglossic language, but one with a strong liturgical strand, which raised questions concerning the sources and extent of religious obligations. The study of secondary school learners, in particular those engaged in a standardised course of language learning, is also much needed in the realms of L2MSS research. Not only does it provide insights into the self-concepts of adolescents, an age viewed as constituting a period of identity-, and thus 'self'-formation, but it also shines an empirical light on learners' interactions with the GCSE, which in the context of Arabic, has eluded scrutiny from a research perspective.

A final, and crucial aspect of this study is the attention paid to the underlying mechanism of the L2MSS—self-discrepancy. Of the limited studies which have analysed this construct in a language learning context, few have interrogated its theorised nature, or the ‘optimal gap’ hypothesis which underlies its postulated motivating power (Section 2.1.2.1.2.3).

6.1.3. Limitations

No study is without its limitations. The inevitable issue of sample size arose from the outset. Whilst data was collected from 265 participants at T1 this fell to 175 students by T2, a drop of almost a third, although analysis showed that attrition was not systematic. The sample size meant restrictions in the number of free parameters available for path modelling which restricted the running of some analyses, such as nested school models. A further complicating factor was the conflation of the classroom and gender variables. Given that most Islamic schools are single-sex, boys and girls were largely clustered within classrooms, potentially confounding analyses comparing genders. Bigger samples are needed, preferably incorporating more schools with mixed-sex classes too, in order to separate out school effects from gender effects, both of which are ripe areas for future research.

A larger sample, and a larger sub-group of HLs would also have facilitated an examination of the effects of proficiency in different dialects upon proficiency outcomes and self-concepts. Research suggests that different dialects exhibit different degrees of overlap with MSA (Kwaik, Saad, Chatzikyriakidis, et al., 2018). Unfortunately, given the limited sample size and diversity of colloquial varieties within it, such statistical comparisons were not possible but would constitute an interesting area for expansion.

A larger sample would also have facilitated examination of the effects of school type on learners’ self-concepts and proficiency. Schools included in the sample were predominantly Muslim faith schools of different types and in different locations around the country. Also included were two non-faith schools, one of which is a well-established private school. The access to resources, training, and student opportunities

will surely differ between such disparate school types. This would constitute an important window onto the effects of the L2 Learning Experience—here operationalised broadly as school environment—on learning and motivational trajectories. A larger sample would also have facilitated the investigation and comparison of classroom- and school-level differences on learners' motivations and proficiency.

The sample is also restricted in terms of diversity of religious background. This, however, is not entirely problematic. As aforementioned, given that the majority of in-school GCSE instruction takes place in Islamic educational institutions, the predominance of Muslim students is not unrepresentative. It does, however, limit the generalisability of these results to non-Muslim, and *ab initio* learners, since as shown, most Muslims GCSE learners begin the course with some degree of proficiency in phonetic decoding. What this also means in practical terms is that the suitability of the GCSE for learners with no background in or exposure to Arabic is hard to assess. It can be said, however, that if learners who do enjoy some prior proficiency are struggling, and initial proficiency is the strongest (and potentially only) predictor of future outcomes in the language, then complete beginners are likely to be at a distinct disadvantage.

Though this study was longitudinal in design, the incorporation of additional timepoints would have facilitated the analyses of longer-term developments in self-concepts and proficiency. This would also reveal whether or not a reciprocal effects model can be validated within a larger timeframe, or at the very least uncover whether self-concepts play any role in promoting proficiency. It is important to bear in mind when interpreting the longitudinal findings of this study that no significant improvements in scores for either the self-concepts or the proficiency measures were recorded between timepoints. Whilst this is an important finding in and of itself, it may explain the detection of particular regression pathways in these models (or lack thereof). Had learners made substantial improvements over the year, a reciprocal effects model may have emerged, hence the need for repeated and longer-term measures.

The tripartite structure of the L2MSS, as originally conceptualised, incorporates the

catch-all category of the L2 Learning Experience (L2LE). To capture this construct to some degree, this study included statistical analyses of the variance between classrooms in terms of self-concepts and proficiency measures. Learners also made mention of teacher effects, and analyses of references to different CS and IS concepts within the interviews revealed frequency concentrations in certain classrooms. The effects of the perception of the course of study also had an impact on learners' CS concepts, as expressed in the interviews. Though insightful, none of these, however, represent systematic measures of the L2LE which are needed in the context of L2MSS research. It is not unreasonable to suspect that, since self-concepts do not appear to drive proficiency, classroom-level factors lie behind learners' initial proficiency levels (since the vast majority had studied Arabic prior to the year under investigation), and likely moderate proficiency gains or lack thereof over the course of learning. Assessments of the learning experience—course content, tasks and resources, teacher, teaching style, peer interactions—represent possibly the most fruitful area for future research in the area of L2 motivation.

One further limitation in this study, it could be argued, is its conceptualisation of Arabic. The language, as previously described, is a spectrum, along which different varieties exist and overlap. The domains of use aligned with each variety were chosen and specified in the survey opening statements. Two things must be acknowledged, however: Firstly, delineating these varieties as discrete, when in reality they do interact, is potentially problematic. It is, however, preferable to overlooking the multiple varieties existent beneath the umbrella of Arabic altogether. Secondly, and relatedly, each variety stretches beyond the domain specified in the surveys. MSA can be integrated into spoken Arabic where necessary, and there is overlap in vocabulary between Classical and MSA. Dialects can also be found written in informal settings, such as on social media. Nevertheless, it remains the case that the three opening statements represent the predominant domains of use for each variety. What should be remembered, and is here being reasserted, is that authentic language use, and actual fluency in the Arabic language require the ability to navigate one's way along this spectrum, upon which these varieties lies and between which borders are porous (Omar, 2011).

The possibility of researcher effects, particularly in the interviews, must also be acknowledged. It is impossible to ascertain to what extent learners' answers to the interview questions, and indeed the way they answered the survey items, were influenced by their perception of who the researcher was, though careful attention was paid in analysing the interview data to any indications of the impact of these views on learners' responses. Note has been made in the instance where this may have been the case (see Section 5.2.1.4.5).

6.1.4. Pedagogical Implications

What do these findings mean for the teaching and learning of Arabic? There are three major implications, two of which apply broadly to the learning of Arabic, and potentially to other languages if the findings of this study are replicated in other language learning contexts. The final implication pertains specifically to the Arabic GCSE.

The emergence of a proficiency-driven model (Section 5.4.1) from the data suggests that emphasis in Arabic language classrooms should be placed on improving learners' actual and perceived proficiency levels in order to promote their self-concepts. Alongside this, learners' IS images appear to be linked to the availability of real-life references to Arabic use, often supplied through personal contact with the language at home or abroad. If the IS has any longer-term role in promoting language learning, educational approaches would be advised to incorporate ways to provide experiences of authentic language use. This may serve to promote learners' Arabic FS-related statements from the realms of the uncertain or hypothetical into the genuinely anticipated and desired.

Findings pertaining to the relationship between classroom-level variables and self-concepts and proficiency measures demonstrate, to some degree, the impact that the learning environment can have on students. Whilst the role of specific components of the learning environment cannot be isolated within the ICC analysis, data from the interviews suggests that the teacher, the level of difficulty of the curriculum, and the compulsory or voluntary nature of the course all contribute substantially to learners' motivations and motivation-related outcomes. This is at once a reassurance and a warning—whilst classroom-level factors can clearly promote motivation, they can also

erode it.

Finally, the lack of progress amongst learners in this sample on proficiency measures derived from GCSE materials, coupled with the finding that learners' self-concept scores were lowest for MSA—the variety taught for the GCSE—calls into question the efficacy and relevance of the GCSE Arabic course. Challenges to the suitability of the GCSE are compounded by the empirical evidence in support of the advantage HLs enjoy and maintain throughout the first year of the course. The mismatch between learners' interests and expectations, as well as underlying parental expectancies related to Islamic schooling (Section 5.2.3.2), presents a significant challenge to the content and delivery of the course. Learners are clearly more interested in the classical and colloquial varieties, neither of which are integrated into the GCSE course, despite (a) the offering of Arabic as part of an Islamic curriculum in many schools in this sample, and (b) claims by the exam board that the material is designed to promote confident and coherent communication with native speakers (Pearson, 2017, p.7) These incongruities need to be addressed at the level of curriculum design and pedagogical practice.

6.1.5. Areas for Future Research

Research inevitably raises as many questions as it answers. This study is no exception. In breaking new ground and focusing on an unexplored language learning terrain, new areas and methods for future research are brought to light and were outlined throughout the discussion chapter. Of these, four stand out as crucial areas for exploration and validation. Firstly, the nature and role of the Past Self requires further investigation. This study revealed the multi-componential nature of the PS—consisting of past learning experiences and experience with authentic language use—and its interactions with both the CS and the IS. Given the novelty of these findings, the existence, nature, and interactions of this component need to be corroborated by research in other contexts. At the other end of the temporal spectrum, the operationalisation of future self-concepts also requires exploration. This study proposes the incorporation of multiple FS components, given the emergence of the theme of contingency and the differing roles of the OTS in relation to shorter- and longer-term FS images. One possible approach, as detailed in Section 5.2.2.3.1 is to ask learners to

provide their own FS concepts, and then to provide details as to whether these are ideal or ought-to driven, and the steps required for them to be realised. Continuing with the theme of self-concepts, the nature of the CS also requires further research. In particular, the role of emotions provides a fruitful area for research. The exact role of affect in the formation and constitution of the CS, rather than as a precedent or corollary to the more cognitive components of this self-concept, needs to be examined.

At the same time as the nature of each of the self-concepts requires focus in future studies, so, too, does their role in motivation. The proficiency-driven model which was observed suggests that self-concepts have little predictive power upon proficiency-based outcomes. If this is indeed the case, then the function of motivational self-concepts requires re-examination. It may be that these constructs assume a predictive function in the longer-term once learners are more established in their course of study (Wigfield & Karpathian, 1991), after which a reciprocal-effects model may be observed. Longitudinal studies incorporating multiple timepoints, both within and across school grades, are required in order for this to be assessed. More pertinent to the discussion on the efficacy of the L2MSS in capturing and explaining motivation is the necessity of further investigating the role of self-discrepancy. This may be facilitated through alternative statistical analyses, such as nested factor models or latent profile analysis, exploring other possible ways of conceptualising and calculating the variable.

An additional, major area for future research is the learning experience (L2LE). Classroom effects were explored in this study via interview analysis and the calculation of ICC coefficients, and were shown to have a substantial role in explaining some of the variance within self-concepts and proficiency levels. However, the L2LE component requires careful operationalisation and measurement in order for as many of the variables existent within the classroom environment and learners' experiences to be identified, captured, and measured. The role of the teacher, resources, teaching method, peer interactions, task-level engagement, course structure and content, all of these require examination as part of the dynamic system at work in the learning environments within which the learner is both agent and object. As well as classroom-level factors, the existence of school-level factors also necessitates examination. What role, for example, does school type or the socio-economic status of learners play in

promoting motivation?

Finally, further research is needed into the replicability of these findings to other Arabic learning contexts. This study suggests that *ab initio* learners may exhibit different self-concepts, particularly since they do not have established past learning experiences to contribute to their current and future self-concepts. In addition, an expansion of the sample to incorporate larger numbers of heritage learners would also facilitate a more detailed examination of the role of different dialects in contributing both to learners' self-concepts and levels of proficiency. This has yet to be done in the realms of L2MSS studies. The investigation of additional Arabic learning contexts, such as mosque courses or optional university courses also offers an area via which to both test and expand the theories proposed in this research, particularly since this would help shine a light on the extent to which perceptions of the course of study interact with perceptions of and progress within the language itself.

To conclude, through the revision of the L2MSS, and the development of new instruments to validate and explore this reconceptualisation, this study presents extensive findings which constitute significant challenges to the components of the model and the interactions between them. These findings also establish empirical grounds upon which to justify the multiglossic conceptualisation of the Arabic language. The study highlights important areas for future research and has implications which extend beyond the realms of theory into the worlds of curriculum design and teaching practice in Arabic classrooms and beyond.

REFERENCES

- Abadzi, H., & Martelli, M. (2014). Efficient Reading for Arab Students: Implications from Neurocognitive Research. *World Summit of Innovation in Education (WISE)*. Doha, Qatar. 5 November. Qatar.
- Abboud-Haggar, S., "Colloquial". (2011). In: Encyclopedia of Arabic Language and Linguistics, Managing Editors Online Edition: Lutz Edzard, Rudolf de Jong. Retrieved from: http://dx.doi.org/10.1163/1570-6699_eall_EALL_SIM_0020.
- Abdelhalim, S., & Alqubayshi, H. (2020). Motivational Orientation and Language Acculturation Experienced by English Speaking Adults Learning Arabic in Saudi Arabia. *Theory and Practice in Language Studies*, 10(9), 1032-1043.
- Abu Bakar, K., Sulaiman, N., Rafeai, Z. (2010) Self-Determination Theory and Motivational Orientations Of Arabic Learners: A Principal Component Analysis. *Journal of Language Studies*, 10(1), 71-86.
- Abu-Rabia, S. (1999). The effect of Arabic vowels on the reading comprehension of second- and sixth- grade native Arab children. *Journal of Psycholinguistic Research*, 28(1), 93-101.
- Ajape, K.O., Mamat, A., & Abdul Azeez, Y. (2015). Students' Motivation and Attitude towards the Learning of Arabic Language: A Case Study of Arabic Students in Nigerian Universities. *International Journal of Economics and Financial Issues*, 5(1), 122-127.
- Al-Batal, M. (2007). Arabic and National Language Educational Policy. *The Modern Language Journal*, 91(2), 268-271.
- Al-Batal, M. (2018). *Arabic as one language: Integrating dialect in the Arabic language curriculum*. Washington.
- Al-Hoorie, A. (2018). The L2 Motivational Self System: A Meta-Analysis. *Studies in Second Language Learning and Teaching*, 8(4), 721-754.
- Al-Hoorie, A., & MacIntyre, P. (2019). *Contemporary Language Motivation Theory: 60 Years Since Gardner and Lambert (1959)* (Psychology of Language Learning and Teaching). Blue Ridge Summit, PA.
- Al Khalil, M. (2011). *Second Language Motivation: Its Relationship to Noticing, Affect, and Production in Task-based Interaction*. (Doctoral Dissertation). Available from ProQuest Dissertations and Theses. (UMI No. 3450687).
- AlAeraini, A. (2004). *Motivational Factors and the Language Proficiency of Learners of Arabic in Selected American Universities*. (Doctoral Dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. 3141460).
- Aladdin, A. (2010). Non-Muslim Malaysian learners of Arabic (NMMLAs): An investigation of their attitudes and motivation towards learning Arabic as a foreign language in multiethnic and multicultural Malaysia. *Procedia, Social and Behavioral Sciences*, 9, 1805-1811.
- Albirini, A. (2014). The Role of the Colloquial Varieties in the Acquisition of the Standard Variety: The Case of Arabic Heritage Speakers. *Foreign Language Annals*, 47(3), 447-463.
- Albirini, A. (2016). *Modern Arabic sociolinguistics: Diglossia, variation, codeswitching, attitudes and identity*. London.
- Alderson, J. C. (1990). Testing reading comprehension skills (Part two): Getting students to talk about taking a reading test (A Pilot Study). *Reading in a Foreign Language*, 7(1), 465-503.

- Alderson, J. C. (2002). Testing proficiency and achievement: principles and practice. In J. A. Coleman, R. Grotjahn, & U. Raatz (Eds.), *University language testing and the C-test*. Bochum: AKS-Verlag. 15-30.
- Alderson, J. C. (2006). *Diagnosing foreign language proficiency: The interface between learning and assessment*. New York: Continuum.
- Alice, M. (2015). *Imputing missing data with R; MICE package 2015*. Retrieved from: <https://datascienceplus.com/imputing-missing-data-with-r-mice-package/>.
- Alhawary, M. (2017). Empirical directions in the future of Arabic second language acquisition research and second language pedagogy. In K. M. Wahba, Z. A. Taha, and L. England (Eds.), *Handbook for Arabic Language Teaching Professionals in the 21st Century II*, (pp. 408-421). London & New York: Routledge.
- Alkhateeb, H. (2014). Reading Anxiety, Classroom Anxiety, Language Motivation, Reader Self-Perception, and Arabic Achievement of Arab-American Students Learning Arabic as a Second Language. *Psychological Reports*, 115(3), 918-931.
- Allen, R. (2004). Perspectives on Arabic teaching and learning. *Modern Language Journal*, 88(2), 275-278.
- Anderson, P. and Suleiman, Y. (2009). Arabic on Campus and Beyond. *British Journal of Middle Eastern Studies*. 36(1), 125–144.
- Anya, U. (2011). Connecting With Communities of Learners and Speakers: Integrative Ideals, Experiences, and Motivations of Successful Black Second Language Learners. *Foreign Language Annals*, 44(3), 441-466.
- Badawī, M. (1973). *Mustawayāt al-‘arabiyya al-mu‘āšira fī Miṣr*. Cairo: Dār al- Ma‘ārif.
- Bale, J. (2010). International Comparative Perspectives on Heritage Language Education Policy Research. *Annual Review of Applied Linguistics*, 30. 42–65.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67(1), 1–48.
- Beatty, P. and Willis, G. (2007) Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, 71(2), 287-311.
- Belnap, K. (2006). A profile of students of Arabic in U.S. universities. In Wahba, K. Taha, Z., and England, L. (Eds) *Handbook for Arabic language teaching professionals in the 21st century*. Routledge. 437-446.
- Belnap, K. (2014) A Profile of Students of Arabic in US Universities. In: K. Wahba., Z. Taha, and L. England, (eds.). *Handbook for Arabic Language Teaching Professionals in the 21st Century*. New Jersey: Routledge, 169-178.
- Bernasek, L. and Canning, J. (2009) Influences on the Teaching of Arabic and Islamic Studies in UK Higher Education: Connections and Disconnections. *Arts and Humanities in Higher Education*, 8 (3), 259-275.
- Blair, R., & Azaz, M. (2019). Self-guide distinctions and episodic memory in L2 motivation: The case of Arabic in the United States. *System*, 81, 192-205.
- Blunch, N. J. (2013). *Introduction to structural equation modeling using IBM SPSS Statistics and Amos* (2nd ed.). London: Sage Publications.
- Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005-2014: Understanding a publication surge and a changing landscape. *System*, 55, 145-157.
- Bouteldjoune, A., McPherron, Paul, Halliday, Laura, & Krassimira, Charkova. (2012). *Motivation in Foreign Language Learning Settings: The Case of Arabic in the USA*. (MSc Thesis). Available from ProQuest Dissertations & Theses Global database. (UMI No. 1529469).
- Boyatzis, R. and Akrivou, K. (2006) The ideal self as the driver of intentional change. *Journal of Management Development*, 25 (7), 624-642.

- British Academy. (2018) *Arabic Mapping Project*. [Website]. Retrieved from: <https://www.thebritishacademy.ac.uk/projects/arabic-mapping-project/>
- Brosh, H. (1988) *The effect of learning Arabic in elementary school on achievement in literary Arabic in the 7th grade*. PhD, Tel Aviv University.
- Brosh, H. (2013). Motivation of American College Students to Study Arabic. *International Journal of Humanities and Social Science*, 3 (19), 27-38.
- Brown, T. A., & Moore, M. T. (2012). *Confirmatory factor analysis*. In R. H. Hoyle (Ed.), *Handbook of structural equation modelling*. New York: Guilford Press. 361-379.
- Buckwalter, T., & Parkinson, D. (2011). *A frequency dictionary of Arabic: Core vocabulary for learners* (Routledge frequency dictionaries). London; New York: Routledge.
- Busch, M., and Turner, J. (1993) Using Likert Scales in L2 Research. *TESOL Quarterly*, 27 (4), 733-739.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS (2nd ed.)*. New York: Routledge.
- Calafato, R. (2020). Learning Arabic in Scandinavia: Motivation, metacognition, and autonomy. *Lingua*, 246,
- Calafato, R., & Tang, F. (2019). The status of Arabic, superdiversity, and language learning motivation among non-Arab expats in the Gulf. *Lingua*, 219, 24-38.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81 –105.
- Campbell, J., Quincy, C., Osserman, J., and Pedersen, O. (2013) Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods & Research*, 42(3), 294-320.
- Carreira, J., (2011). Relationship between motivation for learning EFL and intrinsic motivation for learning in general among Japanese elementary school students. *System*, 39 (1), 90-102.
- Clément, R., Dörnyei, Z. & Noels, K. (1994) Motivation, self-confidence and group cohesion in the foreign language classroom, *Language Learning*, 44(3), 417–448.
- Cohen, A. D. (1998). *Strategies in learning and using a second language*. Longman: Harlow.
- Comanaru, R., & Noels, K. (2009). Self-determination, motivation, and the learning of Chinese as a heritage language. *Canadian Modern Language Review*, 66, 131-158.
- Cranney, A. (1968) *A Comparison of Free-Response and Multiple-Choice Cloze Reading Test*. PhD, University of Minnesota.
- Csizér, K., & Dörnyei, Z. (2005) The internal structure of language learning motivation and its relationship with language choice and learning effort. *Modern Language Journal*, 89(1), 19-36.
- Csizér, K., & Kormos, J. (2009). Learning experiences, selves, and motivated behaviour: a comparative analysis of structural models for Hungarian secondary and university learners of English. In: *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters. 98–119.
- Csizér, K., & Kormos, J. (2014). The ideal L2 self, self-regulatory strategies and autonomous learning: A comparison of different groups of English language learners. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning*. Bristol: Multilingual Matters. 73–88.
- Csizér, K., & Lukács, G. (2010). The comparative analysis of motivation, attitudes and selves: The case of English and German in Hungary. *System*, 38(1), 1-13.

- Cummins, J. (2005). A Proposal for Action: Strategies for Recognizing Heritage Language Competence as a Learning Resource within the Mainstream Classroom. *The Modern Language Journal*, 89 (4), 585-592.
- Demetriou, C., Uzun Ozer, B., & Essau, C. A. (2015). Self-Report Questionnaires. In R. Cautin & S. Lilienfeld (Eds): *The Encyclopedia of Clinical Psychology*. John Wiley & Sons, Inc.
- Department for Education. (2016, April 22). *Community languages saved to ensure diverse curriculum continues*. [Press release]. Retrieved from: <https://www.gov.uk/government/news/community-languages-saved-to-ensure-diverse-curriculum-continues>
- Devlinger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *Methodology*, 13, 31–38.
- Dewaele, J.M., Witney, J., Saito, K. & Dewaele, L., (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language teaching research*, 22(6), 676-697.
- Dickins, J. & Watson, J. C. E. (2006). Arabic Teaching in Britain and Ireland. In: Wahba, K. M., Taha, Z. A., & England, L. (ed.) *Handbook for Arabic language teaching professionals in the 21st century*. New Jersey: Routledge.
- DiStefano, C., Zhu, M., & Mandrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, 14(20), 1–11.
- Djafri, F., & Wimbari, S. (2018). Measuring foreign language anxiety among learners of different foreign languages: In relation to motivation and perception of teacher's behaviors. *Asian-Pacific Journal of Second and Foreign Language Education*, 3(1), 1-15.
- Donato, R., & Davin, K. (2018). The genesis of classroom discursive practices as history-in-person processes. *Language Teaching Research: LTR*, 22 (6), 739-760.
- Donitsa-Schmidt, S., Inbar, O., and Shohamy, E. (2004). The Effects of Teaching Spoken Arabic on Students' Attitudes and Motivation in Israel. *Modern Language Journal*, 88 (2), 217-228.
- Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal*, 78 (3), 273–284.
- Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.
- Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. In Z. Dörnyei (Ed.), *Attitudes, orientations, and motivations in language learning*. Oxford: Blackwell. 3–32.
- Dörnyei, Z. (2005) Motivation and Self-Motivation. In: Z. Dörnyei, Z. (ed.) *The psychology of the language learner: Individual differences in second language acquisition*. New Jersey: Lawrence Erlbaum. 65-118.
- Dörnyei, Z. (2009). The L2 Motivational Self System. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters. 9-42.
- Dörnyei, Z. (2009a). *Questionnaires in second language research* (2nd ed.). London: Routledge.
- Dörnyei, Z. (2010). Researching motivation: From integrativeness to the ideal L2 self. In: S. Hunston & D. Oakey (eds.), *Introducing applied linguistics: Concepts and skills*. London: Routledge, 74-83.

- Dörnyei, Z. (2015). Motivation. In: Z. Dörnyei, and S. Ryan, (eds.) *The psychology of the language learner revisited*. London: Routledge, 72-105.
- Dörnyei, Z. (2019). Psychology and language learning: The past, the present and the future. *Journal for the Psychology of Language Learning*, 1(1), 27-41.
- Dörnyei, Z. (2019a). Towards a better understanding of the L2 Learning Experience, the Cinderella of the L2 Motivational Self System. *Studies in Second Language Learning and Teaching*, IX (1), 19-30.
- Dörnyei, Z. (2020). *Innovations and challenges in language learning motivation*. London: Routledge.
- Dörnyei, Z., & Al-Hoorie, A. (2017). The Motivational Foundation of Learning Languages Other Than Global English: Theoretical Issues and Research Directions. *Modern Language Journal*, 101(3), 455-468.
- Dörnyei, Z. & Chan, L. (2013). Motivation and vision: An analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, 63(3), 437-462.
- Dörnyei, Z., & Kubanyiova, M. (2014). *Motivating learners, motivating teachers: Building vision in language education* (Cambridge language teaching library). Cambridge.
- Dörnyei, Z., MacIntyre, P., & Henry, A. (Eds.) (2015). *Motivational dynamics in language learning*. Bristol: Multilingual Matters.
- Dörnyei, Z., & Ryan, S. (2015). *The Psychology of the Language Learner Revisited*. New York: Routledge.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23(3), 290-325.
- Elman, J. (2003). Development: It's about time. *Developmental Science*, 6(4), 430-433.
- England, L. (2014). Methodology in Arabic Language Teacher Education. In: K. Wahba., Z. Taha, and L. England, (eds.). *Handbook for Arabic Language Teaching Professionals in the 21st Century*. New Jersey: Routledge, 419-436.
- Epskamp, S. (2017). *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output*. [Online]. Retrieved from: <https://github.com/SachaEpskamp/semPlot>
- Eviatar, Z. and Ibrahim, R. (2014). Why is it hard to read Arabic? In: E. Saiegh-Haddad and R. Joshi, (eds.) *Handbook of Arabic Literacy*. Springer: Netherlands, 77-96.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.
- Faraj, R. (2020). Morphosyntactic Transfer from a Spoken Heritage Dialect to the Standard Variety: The Case of Heritage Speakers of Syrian Arabic. *Heritage Language Journal*, 17 (2), 202-233.
- Ferguson, C., 1959. The Arabic Koine. *Language*, 35(4). 616-630.
- Ferrari, G. (2018). *Teaching and Learning Arabic Variation through Vocabulary* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. A13830429).
- Field, A. P. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). London: Sage Publications.
- Field, A. P., Field, Z., & Miles, J. (2012). *Discovering statistics using R*. London: Sage Publications.
- Field, J. (2000). Researching lifelong learning through focus groups. *Journal of Further and Higher Education*, 24(3), 323-335.
- Freire, P. (1985). *Pedagogy of the oppressed* (Pelican books). Harmondsworth: Penguin.

- Gardner, R. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Arnold.
- Gardner, R. (2001). Integrative motivation and second language acquisition. In: Z. Dörnyei and R. Schmidt (eds.). *Motivation and second language acquisition*, Hawai'i: University of Hawai'i Press. 1-20.
- Gardner, R. C. and Lambert, W. (1959). Motivational variables in second language acquisition. *Canadian Journal of Psychology*, 13, 266-72.
- Gardner, R. C., Masgoret, A. M., Tennant, J., and Mihic, L. (2004). Integrative Motivation: Changes During a Year-Long Intermediate-Level Language Course. *Language Learning*, 54 (1), 1-34.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models* (Analytical methods for social research). New York; Cambridge: Cambridge University Press.
- Ghenghesh, P. (2010). The Motivation of Learners of Arabic: Does it Decrease with Age? *Journal of Language Teaching and Research*, 1(3), 235-249.
- Ghanizadeh, A., Eishabadi, N. & Rostami, S. (2016). Motivational dimension of willingness to communicate in L2: The impacts of criterion measure, ideal L2 self, family influence, and attitudes to L2 culture. *International Journal of Research Studies in Education*, 5, 13-24.
- Ghuman, P. (2005) Daughters of Tradition. *The British Psychological Society*, 18, 620-622.
- Gibbs, G. (2012) *Analyzing Qualitative Data*. SAGE Publications, Ltd: London.
- Gough, V., & Calderbank, T. (2019) The teaching and learning of Arabic in British schools: a review of recent research commissioned by The British Council regarding the teaching and learning of Arabic. In: M. Lo (Ed). *The Arabic Classroom: Context, Text and Learners*. Routledge. 274-287.
- Graham, S. (2006). A study of students' metacognitive beliefs about foreign language study and their impact on learning. *Foreign Language Annals*, 39(2), 296-309.
- Graham, S., Courtney, L., Tonkyn, A., & Marinis, T. (2016). Motivational trajectories for early language learning across the primary-secondary school transition. *British Educational Research Journal*, 42(4), 682-702.
- Hae-Young Kim. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38(1), 52-54.
- Hale, G., Stansfield, C., Rock, D., Hicks, M., Butler, F. and; Oller, J. (1989) The relation of multiple-choice cloze items to the Test of English as a Foreign Language. *Language Testing*, 6 (1), 47-76.
- Hale, G., Stansfield, C., Rock, D., Hicks, M., Butler, F., and Oller, J. (1988) *Multiple-choice cloze items and the TOEFL*. TOEFL Research Report No. 26, Princeton, New Jersey: Educational Testing Service.
- Hansen, G. F. (2010). Word recognition in Arabic as a foreign language. *The Modern Language Journal*, 94(4), 567-581.
- Harper, D. (2010) "motive" in *Online Etymology Dictionary*. [Online]. Available from: <http://www.dictionary.com/browse/motive>. [Accessed on: 25 September 2019]
- Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability measures. *Journal of Applied Behavior Analysis*, 10, 103-116.
- Henry, A. (2010). Gender differences in L2 motivation: A reassessment. In: S. A. Davies (Ed.), *Gender gap: Causes, experiences and effects*, New York: Nova Science Publishers. 81-101.

- Henry, A., and C. Cliffordson. (2017). The Impact of Out-of-school Factors on Motivation to Learn English: Self-discrepancies, Beliefs, and Experiences of Self-authenticity. *Applied Linguistics*, 38 (5), 713–736.
- Hessel, G. (2015). From vision to action: Inquiring into the conditions for the motivational capacity of ideal second language selves. *System*, 52, 103-114.
- 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*. Available from: ProQuest Dissertations & Theses Global database. (UMI No. 729003)
- Higgins, E.T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, (94), 319-40.
- Higher Education Funding Council for England (2011) *Strategically important and vulnerable subjects. The HEFCE advisory group's 2010-11 report*. London: HEFCE.
- Hiver, P., Larsen-Freeman, D. (2020). Motivation: It is a relational system. In: Al-Hoorie, A. H., MacIntyre, P. D. (Eds.), *Contemporary language motivation theory: 60 years since Gardner and Lambert (1959)*. Multilingual Matters. 285–303.
- Holes, C. (2004) *Modern Arabic: structures, functions, and varieties*. Washington, D.C.: Georgetown University Press.
- Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14, 332–346.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1- 55.
- Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing*, 19(3), 227-245.
- Husseinali, G. (2006). Who is Studying Arabic and Why? A Survey of Arabic Students' Orientations at a Major University. *Foreign Language Annals*, 39(3). 395–412.
- Husseinali, G. (2012). Arabic Heritage Language Learners: Motivation, expectations, competence, and engagement in learning Arabic. *Journal of the National Council of Less Commonly Taught Languages*, 11, 97-110.
- Ibrahim, R. (2009). The cognitive basis of diglossia in Arabic: Evidence from a repetition priming study within and between languages. *Psychology Research and Behavior Management*, 2, 93-105.
- Ibrahim, Z. and Allam, J. (2006). Arabic Learners and Heritage Students Redefined: Present and Future. In Wahba, K., Taha, Z., and England, L. (eds.) *Handbook for Arabic Language Teaching Professionals in the 21st Century*. New Jersey: Routledge. 437- 446.
- Islam, M., Lamb, M., and Chambers, G. (2013). The L2 Motivational Self System and National Interest: A Pakistani perspective. *System*, 41(2), 231-244.
- Jaspal, R., & Coyle, A. (2010). "Arabic is the language of the Muslims-that's how it was supposed to be": Exploring language and religious identity through reflective accounts from young British-born South Asians. *Mental Health, Religion & Culture*, 13(1), 17-36.
- Jonz, J. (1976). Improving on the basic egg: the M-C cloze. *Language Learning*, 26 (2), 255-65.
- Kaushik, V., & Walsh, C. (2019). Pragmatism as a Research Paradigm and Its Implications for Social Work Research. *Social Sciences*, 8(9), 255.

- Kendall, E. (2014). *The Challenges Facing the Creation of Arabic Expertise in the UK*. [Lecture]. Held on 27 May 2014, British Academy, London.
- Khamis-Dakwar, R., & Froud, K. (2007). Lexical processing in two language varieties: An event-related brain potential study of Arabic native speakers. In: M. Mughazy (ed.), *Perspectives on Arabic linguistics XX*. Amsterdam & Philadelphia: John Benjamins. 153–68.
- Kikuchi, K. (2017). Reexamining demotivators and motivators: A longitudinal study of Japanese freshmen's dynamic system in an EFL context. *Innovation in Language Learning and Teaching* 11, 128–145.
- Kim, H.-Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38(1), 52-54.
- Kim, Y.-K., & Kim, T.-Y. (2011). The Effect of Korean Secondary School Students' Perceptual Learning Styles and Ideal L2 Self on Motivated L2 Behavior and English Proficiency. *Korean Journal of English Language and Linguistics*, 11(1), 21-42.
- Klare, G., Sinakiaio, H., and Stulran, L. (1972) The Cloze procedure: a convenient readability test for training materials and translations. *International Review of Applied Psychology*. 21 (3), 77-106.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York: Guildford Press.
- Kong, J. H., Han, J. E., Kim, S., Park, H., Kim, Y. S., & Park, H. (2018). L2 motivational self system, international posture and competitiveness of Korean CTL and LCTL college learners: A structural equation modeling approach. *System*, 72, 178-189.
- Kormos, J., Kiddle, T., & Csizér, K. (2011) Goals, attitudes and self-related beliefs in second language learning motivation: an interactive model of language learning motivation. *Applied Linguistics*, 32(5), 495-516.
- Kwaik, K., Saad, M., Chatzikyriakidis, S., & Dobnika, S. (2018). A Lexical Distance Study of Arabic Dialects. *Procedia Computer Science*, 142, 2-13.
- Lamb, M. (2009). Situating the L2 self: two Indonesian school learners of English. In: Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters. 229-247.
- Lamb, M. (2012). A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning*, 62(4), 997–1023.
- Lantolf, J.P. and Swain, M. (2019). Perezhivanie: The cognitive–emotional dialectic within the social situation of development. In: A. Al-Hoorie and P.D. MacIntyre (eds) *Contemporary Language Motivation Theory: 60 Years Since Gardner and Lambert (1959)*. Bristol: Multilingual Matters.
- Lanvers, U. (2016). Lots of Selves, some rebellious: Developing the Self Discrepancy Model for Language Learners. *System*, 60, 79-92.
- Lee, Y. S., & Ahn, K. J. (2013). English learning motivation and English academic achievement of Korean elementary school students: The effects of L2 selves, international posture, and family encouragement. *Modern English Education*, 14(1), 127-152
- Lepp-Kaethler, E., & Dörnyei, Z. (2013). The role of sacred texts in enhancing motivation and living the vision in second language acquisition. In Wong, M. S., Kristjánsson, C., & Dörnyei, Z. (Eds.), *Christian faith and English language teaching and learning: Research on the interrelationship of religion and ELT*. New York: Routledge. 171-188.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: Sage Publications.

- MacIntyre, P., & Gregersen, T. (2012). Emotions that facilitate language learning: The positive- broadening power of the imagination. *Studies in Second Language Learning and Teaching*, 2, 193–213.
- MacIntyre, P., Mackinnon, S. & Clément, R. (2009) Toward the development of a scale to assess possible selves as a source of language learning motivation. In: Z. Dörnyei and E. Ushioda (eds) *Motivation, Language Identity and the L2 Self*. Bristol: Multilingual Matters. 193–214.
- Macintyre, P., & Vincze, L. (2017). Positive and negative emotions underlie motivation for L2 learning. *Studies in Second Language Learning and Teaching*, 7(1), 61-88.
- Macmillan, N. A. (2002). Signal Detection Theory. In: H. Pashler, & J. Wixted (Eds.), *Stevens' Handbook of Experimental Psychology*. New York: John Wiley and Sons. 55–91.
- Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, 110, 63-73.
- Mango, O. (2011). Arabic Heritage Language Schools in the US. Heritage Language Briefs. Center for Applied Linguistics. [Online]. Retrieved from: <http://www.cal.org/heritage/pdfs/briefs/arabic-heritage-language-schools-in-the-us.pdf>.
- Markus, H. and Nurius, P. (1986) Possible selves. *American Psychologist*, 41, 954-69
- Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, 34, 155–167.
- Marsh, H. W., Pekrun, R., Lichtenfeld, S., Guo, J., Arens, A. K., & Murayama, K. (2016). Breaking the double-edged sword of effort/trying hard: Developmental equilibrium and longitudinal relations among effort, achievement, and academic self-concept. *Developmental Psychology*, 52(8), 1273–1290.
- McMillan, J., & Hearn, J. (2008). Student Self-Assessment: The Key to Stronger Student Motivation and Higher Achievement. *Educational Horizons*, 87(1), 40-49.
- Meara, P. (2010). *EFL vocabulary tests* (2nd ed.). Centre for Applied Language Studies.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142–154.
- Mendoza, A., & Phung, H. (2019). Motivation to learn languages other than English: A critical research synthesis. *Foreign Language Annals*, 52(1), 121-140.
- Meshkat, M. & Saeb, F. (2014). High-school Students' Beliefs about Learning English and Arabic. *International Journal of Applied Linguistics & English Literature*, 3(3), 211-216.
- Miles, M., Huberman, A., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook* (Third ed.). Los Angeles: London.
- Milton, J. (2006). French as a foreign language and the Common European Framework of Reference for languages.[Online]. Retrieved from: <https://www.llas.ac.uk/resources/paper/2715>
- Milton, J. (2006a). Language lite? Learning French vocabulary in school. *Journal of French Language Studies*, 16(2).
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. In: Bartning, I., Martin, M., and Vedder, I., (Eds.) *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*. EuroSLA.

- Mohee, S. (2011). *Young British South Asian Muslim Women: Identities and Marriage*. [Doctoral dissertation]. Available from: ProQuest Dissertations & Theses Global database. (UMI No. 568308).
- Moran-Ellis, J., Alexander, V., Cronin, A., Dickinson, M., Fielding, J., Sleney, J., & Thomas, H. (2006). Triangulation and integration: Processes, claims and implications. *Qualitative Research*, 6(1), 45-59.
- Moskovsky, C., Assulaimani, T., Racheva, S., & Harkins, J. (2016). The L2 Motivational Self System and L2 Achievement: A Study of Saudi EFL Learners. *Modern Language Journal*, 100(3), 641-654.
- Murphy, V. (2014). *Second Language Learning in the Early School Years*. Oxford: Oxford University Press.
- Nakamura, T. (2019). Understanding motivation for learning languages other than English: life domains of L2 self. *System* 82, 111-121.
- Newberg, A. (2013). *The Spiritual Brain: Science and Religious Experience*. [Audiobook]. The Great Courses: Audible. Retrieved from: https://www.audible.co.uk/pd/The-Spiritual-Brain-Science-and-Religious-Experience-Audiobook/B00DJ6QJSY?qid=1617111596&sr=1-2&ref=a_search_c3_lProduct_1_2&pf_rd_p=c6e316b8-14da-418d-8f91-b3cad83c5183&pf_rd_r=WVFN7Z1S3V8BWTWJ9KZR
- Nielsen, H. (2011) Second Language Teaching. In: Edzard, L. and de Jong, R. (eds.). *Encyclopedia of Arabic Language and Linguistics*, Brill. [Online]. Retrieved from: http://ezproxy-prd.bodleian.ox.ac.uk:2134/entries/encyclopedia-of-arabic-language-and-linguistics/second-language-teaching-EALL_COM_0297
- Norman, C., & Aron, A. (2003). Aspects of future self that predict motivation to achieve or avoid it. *Journal of Experimental Social Psychology*, 39, 500-507.
- Norton, B. (2020). Motivation, identity and investment: A journey with Robert Gardner. In A. H. Al-Hoorie & P. D. MacIntyre (Eds.), *Contemporary language motivation theory: 60 years since Gardner and Lambert (1959)*. Multilingual Matters. 153-168.
- Nunan, D., Aronson, J., & Bankhead, C. (2018). Catalogue of bias: attrition bias. *BMJ evidence-based medicine*, 23(1), 21-22.
- Office for National Statistics; National Records of Scotland; Northern Ireland Statistics and Research Agency (2017). *2011 Census aggregate data*. UK Data Service (Edition: February 2017).
- Ofqual. (2020). Relationships between grade distributions for students taking different combinations of reformed GCSE subjects. [Online]. Retrieved from: <https://analytics.ofqual.gov.uk/apps/GCSE/9to1/>
- Olsen, M. (2018). What and Who in the L2 Motivational Self System: The Interlinked Roles of Target Language and Heritage Learner Status. *Canadian Modern Language Review*, 74 (2), 279-301.
- Omar, N. (2011). Diglossia. In: Edzard, L. and de Jong, R. (eds.). *Encyclopedia of Arabic Language and Linguistics*, Brill. [Online]. Available from: http://ezproxy-prd.bodleian.ox.ac.uk:2134/entries/encyclopedia-of-arabic-language-and-linguistics/second-language-teaching-EALL_COM_0297 [Accessed on: 25 September 2019].
- Oyserman, D., & James, L. (2009). Possible selves: From content to process. In K. Markman, W. M. P. Klein, & A. Suhr (Eds.), *The handbook of imagination and mental stimulation*. New York: Psychology Press. 373-394.
- Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System*, 38, 467-479.

- Papi, M., & Abdollahzadeh, E. (2012). Teacher Motivational Practice, Student Motivation, and Possible L2 Selves: An Examination in the Iranian EFL Context. *Language Learning*, 62(2), 571-594.
- Papi, M., Bondarenko, A., Mansouri, S., & Feng, L. (2019). Rethinking L2 Motivation Research: The 2 × 2 Model of L2 Self-Guides. *Studies in Second Language Acquisition*, 41(2), 337-361.
- Pearson Edexcel (2017). GCSE (9-1) Arabic: Specification. Pearson Edexcel Limited.
- Pearson Edexcel, (2019). *Examiners' Report: Principal Examiner Feedback. Summer 2019*. Pearson Edexcel Limited.
- Pekrun R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315-341.
- Pekrun, R., Lichtenfeld, S., Marsh, H., Murayama, K., & Goetz, T. (2017). Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects. *Child Development*, 88(5), 1653-1670.
- Qafisheh, H. A. (1972). *From Gulf Arabic into Modern Standard Arabic: A pilot study*. The 6th ACTFL annual meeting. Atlanta, Georgia.
- Qatar Foundation International. (2021). Infographic: Dialects of the Arab World. [Website]. Retrieved from: <https://www.qfi.org/blog/infographic-dialects-arab-world/>
- Ramezanzadeh, A. (2015). *Exploring GCSE Arabic: An analysis of heritage and non-heritage language learner performance*. MSc Applied Linguistics and Second Language Acquisition, University of Oxford.
- Ramezanzadeh, A. (2016). One Size Fits All? An analysis of heritage and non-heritage language learner performance in GCSE Arabic. [Report]. British Council. Retrieved from: https://www.britishcouncil.org/sites/default/files/an_analysis_of_heritage_and_non-heritage_language_learner_performance_in_gcse_arabic_28conducted_by_anna-maria_ramezanzadeh_2016.pdf29.pdf
- Ramezanzadeh, A. (In preparation). *Analysis of GCSE and A Level Arabic*. In: Kohl, Katrin, et al. *Report on MFL assessment at GCSE and A level in smaller-cohort languages*. Oxford.
- Random House. (2012). Motivation. In: *Random House Dictionary*. New York: Random House, Inc.
- Resto, J. (2013). What is Arabic? In: J. Owens (ed.). *The Oxford Handbook of Arabic Linguistics*. Oxford: OUP, 433-450.
- Ricks, R. (2015). *The Development of Frequency-based Assessments of Vocabulary Breadth and Depth for L2 Arabic*, Georgetown University. ProQuest Dissertations and Theses.
- Rosowsky, A. (2006). The role of liturgical literacy in UK Muslim communities. In T. Omoniyi, & J.A. Fishman (Eds.), *Explorations in the sociology of language and religion*. Amsterdam: John Benjamins. 309-324.
- Rosowsky, A. (2007). Qur'anic literacy: Its central role in the life of UK Muslim communities. *NALDIC Quarterly*, 5(2), 44.
- Rosowsky, A. (2008). *Heavenly Readings: Liturgical Literacy in a Multilingual Context*. Bristol: Multilingual Matters.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://www.jstatsoft.org/v48/i02/>.

- Roulston, K. (2010). Considering quality in qualitative interviewing. *Qualitative research*, 10 (2), 199–228.
- Ryan, R. and Deci, E. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25, 54–67.
- Ryan, S. (2009). Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self*. Clevedon, UK: Multilingual Matters. 120–143.
- Saito, K., Dewaele, J., Abe, M., & In'nami, Y. (2018). Motivation, Emotion, Learning Experience, and Second Language Comprehensibility Development in Classroom Settings: A Cross-Sectional and Longitudinal Study. *Language Learning*, 68(3), 709–743.
- Saldaña, J. (2012). *The coding manual for qualitative researchers* (2nd ed.). London.
- Sampasivam, S., & Clément, R. (2014). The Dynamics of Second Language Confidence: Contact and Interaction. In: S. Mercer and M. Williams (Eds.) *Multiple Perspectives on the Self in SLA*. Bristol: Multilingual Matters. 23–40.
- Savin-Barden, M., and Howell Major, C. (2013). *Qualitative Research: The essential guide to research and practice*. Routledge.
- Scholz, G., and Scholz, C. (1981). *Multiple-choice cloze tests of EST discourse: An exploration*. Paper presented at the 15th annual TESOL convention. Detroit.
- Schumann, J. (1997). *The neurobiology of affect in language learning*. Oxford: Blackwell.
- Scott-Baumann, A. and Contractor, S. (2012). *Arabic Language and the Study of Islam: Who studies Arabic and why and how can these skills be used at university and beyond?* [Research Report]. LLAS Centre for Languages, Linguistics and Area Studies, University of Southampton.
- Semaan, G., & Yamazaki, K. (2015). The Relationship Between Global Competence and Language Learning Motivation: An Empirical Study in Critical Language Classrooms. *Foreign Language Annals*, 48(3), 511–520.
- Seymour-Jorn, C. (2004). Arabic language learning among Arab immigrants in Milwaukee, Wisconsin: A study of attitudes and motivations. *Journal of Muslim Minority Affairs*, 24(1), 109–122.
- Shapka, J., Keating, D. (2005). Structure and Change in Self-Concept during Adolescence. *Canadian Journal of Behavioural Science*, 37 (2), 83–96.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4).
- Stevens, P. (2006). Is Spanish Really So Easy? Is Arabic Really So Hard?: Perceived Difficulty in Learning Arabic as a Second Language. In: K. Wahba., Z. Taha, and L. England, (eds.). *Handbook for Arabic Language Teaching Professionals in the 21st Century*. New Jersey: Routledge. 35–66.
- Stubbs, J., and Tucker, G. (1974). The Cloze Test as a Measure of English Proficiency. *Modern Language Journal*, 58 (5–6), 239–241.
- Syed, M., and Nelson, S. (2015). Guidelines for Establishing Reliability When Coding Narrative Data. *Emerging Adulthood*, 3 (6), 375–387.
- Tabachnik, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Harlow: Pearson Education.
- Taguchi, T., Magid, M. and Papi, M. (2009). The L2 motivational self system amongst Chinese, Japanese and Iranian learners of English: A comparative study. In: Z. Dörnyei and E. Ushioda (eds.). *Motivation, language identity and the L2 self*. Clevedon: Multilingual Matters, 66–97.

- Taha, T. (2007). Arabic as "A Critical-Need" Foreign Language in Post-9/11 Era: A Study of Students' Attitudes and Motivation. *Journal of Instructional Psychology*, 34(3), 150-160.
- Taylor, S., Bogdan, R., & DeVault, M. (2016). *Introduction to qualitative research methods : A guidebook and resource*. (Fourth ed., Ebook central). Hoboken, New Jersey.
- Taylor, F., & Marsden, E. J. (2014). Perceptions, attitudes, and choosing to study foreign languages in England: an experimental intervention. *The Modern Language Journal*, 98(4), 902-920.
- Taylor, W. (1956). Recent Developments in the use of Cloze procedure. *Journalism Quarterly*, 33, 42-48.
- Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, 39, 681-709.
- Temples, A. (2010). Heritage Motivation, Identity, and the Desire to Learn Arabic in U.S. Early Adolescents. *Journal of the National Council of Less Commonly Taught Languages*, 9, 103-132.
- Thompson, A., S. (2017). Don't tell me what to do! The anti-ought-to self and language learning motivation. *System*, 67, 38-49.
- Thorsen, C., Henry, A., & Cliffordson, C. (2020). The case of a missing person? The current L2 self and the L2 Motivational Self System. *International Journal of Bilingual Education and Bilingualism*, 23(5), 584-600.
- Tinsley, T. (2015). *The teaching of Arabic language and culture in UK schools*. Alcantara Communications. [Online]. Available from: https://www.britishcouncil.org/sites/default/files/arabic_report_2015.pdf
- Tinsley, T. (2019). *Language Trends*. British Council. Available from <https://www.britishcouncil.org/research-policy-insight/research-reports/language-trends-2019>
- Tinsley, T. (2020). *Coronavirus and language GCSEs, 2020*. [Blog post]. Retrieved from: <http://www.alcantaracomms.com/coronavirus-and-languages-gcses-2020/>
- Tinsley, T. and Board, K. (2013). *Languages for the Future: Which languages the UK needs most and why*. British Council.
- Tinsley, T. and Board, K. (2017). *Languages for the Future: Which languages the UK needs most and why*. British Council.
- Toeimah, R. (1978). *The Use of Cloze to Measure the Proficiency of Students of Arabic as a Second Language in some universities in the United States*. PhD, University of Minnesota.
- Trentman, E., & Shiri, S. (2020). The Mutual Intelligibility of Arabic Dialects: Implications for the Language Classroom. *Critical Multilingualism Studies*, 8 (1), 104-134.
- US Department of State (n.d.) *Languages*. [Online]. Retrieved from: <https://www.state.gov/m/fsi/sls/orgoverview/languages/>
- Ushioda, E. (2009). A person-in-context relational view of emergent motivation, self and identity. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters. 215-228.
- Ushioda, E. (2013). Christian faith, motivation and L2 learning: Personal, social and research perspectives. In: M. S. Wong, C. Kristjansson, & Z. Dörnyei (Eds.), *Christian faith and English language teaching and learning: Research on the interrelationship of religion and ELT*. New York: Routledge/Taylor & Francis. 223-229.

- Ushioda, E. (2014) Motivational Perspectives on the Self in SLA: A Developmental View. In S. Mercer & M. Williams (Eds.), *Multiple perspectives on the self in SLA*. Bristol: Multilingual Matters. 127-141.
- Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. Peyton, D. Ranard, Q S. McGinnis (eds.), *Heritage languages in America: Preserving a national resource*. Washington, DC: Center for Applied Linguistics and Delta Systems. 37-77.
- Van Buuren, Stef, Groothuis-Oudshoorn, Catharina Gerarda Maria, & Faculty of Behavioural, Management Social Sciences. (2011). Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1548-7660.
- Versteegh, C. (2014). The Arabic language (Second ed.). Edinburgh.
- Visonà, M., & Plonsky, L. (2019). Arabic as a heritage language: A scoping review. *International Journal of Bilingualism*, 24(4), 599-615.
- Vygotsky, L.S. (1971). *The Psychology of Art*. Cambridge, MA: MIT Press.
- Wahba, K. (2006). Arabic Language Use and the Educated Language Use. In: K. Wahba., Z. Taha, and L. England, (eds.). *Handbook for Arabic Language Teaching Professionals in the 21st Century*. New Jersey: Routledge, 139-176.
- Watt, J. M., and Fairfield, S. L. (2008). Religious and Sacred Literacies. In: B. Spolsky, & F. Hult, (Eds.). *The Handbook of Educational Linguistics* (Blackwell handbooks in linguistics). Malden, MA; Oxford: Blackwell. 355-366.
- Weiner, B. (1986). *An attributional theory of motivation and emotion* (Springer series in social psychology). New York; London: Springer-Verlag.
- Wigfield, A., & Karpathian, M. (1991). Who Am I and What Can I Do? Children's Self-Concepts and Motivation in Achievement Situations. *Educational Psychologist*, 26(3-4), 233-261.
- Williams, M., Burden, R., & Lanvers, U. (2002). 'French is the Language of Love and Stuff': Student perceptions of issues related to motivation in learning a foreign language. *British Educational Research Journal*, 28(4), 503-528.
- Winke, P., and Weger-Guntharp, H. (2006). Why students in the U.S. are learning Arabic: A study of motivation at the college-level. *Journal of the National Council of Less Commonly Taught Languages*, 3, 7-33.
- Wolak, M.E., Fairbairn, D.J., & Paulsen, Y.R. (2012). Guidelines for Estimating Repeatability. *Methods in Ecology and Evolution*, 3(1), 129-137.
- Yang, J. S., & Kim, T. Y. (2011). The L2 Motivational Self-System and perceptual learning styles of Chinese, Japanese, Korean, and Swedish students. *English teaching*, 66(1), 141-162.
- You, C., & Dörnyei, Z. (2016). Language Learning Motivation in China: Results of a Large-Scale Stratified Survey. *Applied Linguistics*, 37(4), 495-519.
- You, C., Dörnyei, Z., & Csizér, K. (2016). Motivation, Vision, and Gender: A Survey of Learners of English in China. *Language Learning*, 66(1), 94-123.
- Younes, M. (2006). Integrating the colloquial with fuṣḥā in the Arabic as a foreign language classroom. In K.Wahba, Z.Taha, and L. England (Eds.), *Handbook for Arabic language teaching professionals in the 21st century*, (pp. 157–166). Mahwah, NJ: Lawrence Erlbaum.
- Zabarah, H. (2016). College-Level Arabic Heritage Learners: Do they belong in Separate Classrooms? *Journal of the National Council of Less Commonly Taught Languages*, 18, 93-120.
- Zentner, Marcel, & Renaud, Olivier. (2007). Origins of adolescents' ideal self: An intergenerational perspective. *Journal of Personality and Social Psychology*, 92(3), 557-74.

APPENDICES

A - LITERATURE REVIEW

A.1. Word Frequency Band Analysis

(Excerpt take from report authored by the researcher: Ramezanzadeh, A. (In preparation). *Analysis of GCSE and A Level Arabic*. In: Kohl, Katrin, et al. *Report on MFL assessment at GCSE and A level in smaller-cohort languages*. Oxford.)

Vocabulary

There is no consensus on exactly how language GCSEs and A Levels align with the Common European Framework of Reference (CEFR) bands, however Milton (2006) suggests that a high grade at GCSE and A Level are roughly equal to the CEFR proficiency levels of A2 and B2 respectively. The framework does not specify vocabulary size for each of these levels, however researchers estimate approximately 2,000 words are needed for A2 and up to 4,000 for B2 (Milton, 2010). This will of course vary based on language, however since there are no Arabic-specific estimates these brackets will be used for the purposes of subsequent vocabulary analysis.

The vocabulary items in the list provided in the GCSE specification, and the items used in the papers themselves were cross-referenced with Buckwalter and Parkinson's *A Frequency Dictionary of Arabic: Core Vocabulary for Learners* (2011) which lists the top 5,000 most frequent words in Modern Standard Arabic.

The Specification for the reformed GCSE states that 'texts for individual questions within the assessment use high-frequency language' (2017, p. 36). However, a breakdown of the verb section of the Arabic vocabulary list for the GCSE taken from the reformed specification by word frequency band produced the following results.

This indicates that over 50% of the vocabulary items included in the list provided are above the 2,000 recommended for this level of proficiency (see Table A.1.1). Of these, 27% of the words provided are beyond the 5,000 remit which is beyond even the recommended level for a high grade at A Level. Whilst it is clear that some subject-specific low frequency vocabulary will necessarily be included in the lists in order to facilitate comprehension of the themes of the GCSE, it is doubtful that this can explain more than half of the items being beyond the expected scope of learners according to the CEFR.

Table A.1.1. Reformed Verb Vocabulary List by Frequency Band			
Word Frequency Band	Words	%	%
0-1000	80	26.1	46.7
1000-2000	63	20.6	
2000-3000	40	13.1	53.3
3000-4000	27	8.8	
4000-5000	14	4.6	
5000 +	82	26.8	
Total	306	100%	

Table A.1.2. Legacy Vocabulary List by Frequency Band			
Word Frequency Band	Words	%	%
0-1000	518	27.4	43.8
1000-2000	310	16.4	
2000-3000	220	11.6	56.2
3000-4000	139	7.4	
4000-5000	103	5.4	
5000+	600	31.7	
Total	1890	100%	

Whilst these figures refer only to the verb section of the reformed vocabulary list, a similar analysis was conducted on the entire legacy list, which totalled 1890 words. This resulted in comparable findings (see Table A.1.2).

The inclusion of low frequency words in texts is, to some degree, unavoidable, and it is possible in some instances to infer the meaning of such items from context, and root/morphological pattern (although this requires the development of considerable linguistic skills). However, the sheer volume of low frequency words, those beyond the CEFR-based brackets and those beyond the 5,000 mark, presents a huge challenge to learners who will need to interact with these items, in the context of the exams, without recourse to a dictionary, and within a limited timeframe.

A.2. Arabic Motivation Studies Summary Table

	Author(s)	Year	Pub. Type	Title	Country	Pop.	Sample Size	Motivation Framework	Analysis	Data Collection Methods	RQs/ Research Focus	Key Findings (Motivation)
1	AlAeraini	2004	PhD Thesis	Motivational Factors and the Language Proficiency of Learners of Arabic in Selected American Universities	USA	University students	266	Mixed (SDT? Gardner?) Instrument adapted from Ryu-Yang (2001)	Quant	Questionnaire	1. Is there a significant relationship between EM and self-reported language proficiency? 2. Is there a significant relationship between IM and self-reported language proficiency? 3. What levels of proficiency do American learners of the Arabic language in American universities attain? 4. What is the relation between the learners' EM and their gender and educational level? 5. What is the relation between the learners' IM and their gender and educational level? 6. Are American learners of Arabic language more motivated by Extrinsic or Intrinsic Factors?	Intrinsic motivation (IM) factors more motivating than extrinsic motivation (EM) factors; Also found relationships between language proficiency and both extrinsic and intrinsic motivation factors; no correlation between extrinsic motivation, intrinsic motivation and gender; a significant weak correlation between extrinsic and intrinsic motivations and language proficiency; Negative weak significant correlation between EM and educational level; No significant correlation between IM and educational level.
2	Donitsa-Schmidt et al.	2004	Journal Article	The Effects of Teaching Spoken Arabic on Students' Attitudes and Motivation in Israel	Israel	Elementary school students (4-6 th grade)	692	Gardner (Integrative; Instrumental)	Quant	Questionnaires – one experimental group (studying spoken Arabic); one control group (not studying Arabic)	1. Does the teaching of spoken Arabic at a younger age affect student attitudes towards the Arabic language and culture and their motivation to study the language? 2. What are the attitudes of the students' parents towards the teaching of Arabic?	students who study spoken Arabic (experimental group) hold more positive attitudes towards the Arabic language, its culture, and speakers, than control group; Experimental group also claim to be more motivated to study the language;

											3. Which variables best predict the students' motivation to study Arabic in the future?	parents' attitudes were a predictor of students' motivation to study Arabic; learners' satisfaction with Arabic study program best predicted students' motivation.
3	Seymour-Jorn	2004	Journal Article	Arabic language learning among Arab immigrants in Milwaukee, Wisconsin: a study of attitudes and motivations	USA	School students, university students, community members	15 university students	Not specified	Mixed	Questionnaire; Interviews; Observation	what role does the teaching and learning of Arabic have in these families for helping members to feel that they are transmitting their cultural and religious heritage? Does emphasis on Arabic language learning have a mostly religious or cultural dimension? Does learning Arabic help Arab-Americans to situate themselves with regard to the global multicultural, or to family and others in the Arab world?	interest in Arabic stems from feeling of community members that young Muslims have a responsibility to read the Qur'an and understand it. Uni students: 50% of respondents referenced desire to read the Qur'an in Arabic and make their own interpretations of the text; 50% of respondents cited desire to improve reading and writing skills.
4	Husseinali	2006	Journal Article	Who is Studying Arabic and Why? A Survey of Arabic Students' Orientations at a Major University	USA	University students	120	Not specified (instrument adapted from Wen 1997)	Quant	Questionnaire	1. What are the most important orientations of AFL learners as a group? 2. What are the most important orientations for heritage learners? For non- heritage learners? 3. Which orientations would make consistent factors of orientations in AFL? 4. Are there significant orientation differences between groups of learners based on heritage background, level, and gender?	HLs significantly more motivated than NHLs to study AFL for identification reasons; NHLs significantly more motivated to study Arabic for instrumental reasons; No significant differences between groups on travel and culture orientations; No significant differences between male and female learners.
5	Winke & Weger-	2006	Journal	Why students	USA	College/	326	Processing	Quant	Arabic	1) What are the demographics of	Ten primary motivation categories

	Guntharp		Article	in the U.S. are learning Arabic: A study of motivation at the college-level.		university students		model of motivation	(motivation qual)	proficiency test (reading and listening); Questionnaire (one open-ended question about motivation)	students studying Arabic at U.S. colleges and universities? 2) What are their primary motivations for learning Arabic? 3) Are motivations for learning Arabic related to proficiency levels?	emerged from the data; Learners at higher levels of study/obtained higher levels of proficiency expressed different motivations from those at lower levels - corresponds to a processing conceptualization of motivation; Largest category was 'employment/job' (26%); HLs had motivations distinct from NHLs, and from those with other L1 backgrounds; Motivations varied according to proficiency and L1 background.
6	Taha	2007	Journal Article	Arabic as "A Critical-Need" Foreign Language in Post-9/11 Era: A Study of Students' Attitudes and Motivation	USA	Historically Black University (HBCU)	142 (34 PG; 108 UG)	Not specified (Integrative; Instrumental)	Quant	Questionnaire (inc open ended Q)	Explore attitudes, motivations, and perceptions of students regarding learning/ teaching of AFL in post-9/11 era	Learning primarily for utilitarian proposes: 61.2% Travel/tourism 84.5% Linguistic/cultural asset
7	Abu Bakar et al.	2010	Journal Article	Self-Determination Theory And Motivational Orientations Of Arabic Learners: A Principal Component Analysis	Malaysia	University students	228	Self-Determination Theory	Quant	Questionnaire	Does the 7-factor structure of motivational orientation, which characterized Anglo-French learners of English and French exist among Muslim learners of Arabic?	4 categories (Intrinsic Motivation-Knowledge, Intrinsic Motivation-Accomplishment, Extrinsic-Motivation Identified Regulation, and Amotivation) emerged as reliable and valid; Intrinsic-Knowledge accounted for 34% of the variance; Religious motivation emerged as a new statistical subdimension of motivational orientation - explained 5.43% of the variance - consists of items with both intrinsic and extrinsic labels.

8	Aladdin	2010	Journal Article	Non-Muslim Malaysian learners of Arabic (NMMLAs): an investigation of their attitudes and motivation towards learning Arabic as a foreign language in multiethnic and multicultural Malaysia	Malaysia	University students (Non-Muslim Malaysian (NMMLA))	207	Gardner (inc AMTB)	Quant	Questionnaire	investigating the motivational orientations among the NMMLAs in learning Arabic as a foreign language.	Findings reveal that the NMMLAs' held a high positive attitudes towards foreign languages, and held a moderate positive attitudes towards Arabic language and its native speakers; Respondents instrumentally motivated to learn Arabic language - need to fulfil the university requirement had highest ranking; Intrinsic and integrative motivation came second; Respondents perceived Arabic is not an easy language to learn
9	Ghenghesh	2010	Journal Article	The Motivation of Learners of Arabic: Does it Decrease with Age?	Libya	Senior international high school students (and teachers)	144	Unclear	Mixed	Questionnaires; Interviews	'(a) the extent in which various factors affect students' motivation and achievement in second language learning particularly as they enter the senior high school; (b) to carry out a cross-comparison of the results obtained by the questionnaire with those of the interviews and (c) to obtain additional information on how students' motivation can be enhanced.'	Older learners score significantly lower on motivational scales, supported by interview data; Role of the teacher was seen fundamental in determining the attitude to the language and in supplying motivation.
10	Temples	2010	Journal Article	Heritage Motivation, Identity, and the Desire to Learn Arabic in U.S. Early Adolescents.	USA	Secondary school learners (grades 6-8) (and their parents)	80	Unclear	Qual	Questionnaires, interviews, classroom observations of students and parents	Validation of existence of, and examination of differences between proficiency and motivation of 3 groups of learners: learners of Arabic descent (HLL), non-Arab Muslims	School requirement is highest motivation (71.3%), followed by parents (45%). But motivations vary between three groups: FLL group showed different profiles to others. No religious motivations.

											(RHL), and learners with other language and cultural backgrounds (FLL).	
11	Bouteldjouné	2012	MSc Thesis	Motivation in Foreign Language Learning Settings: The Case of Arabic in the USA	USA	College students; ESL instructors	13	Not specified (adapted Husseinali 2006 survey – Instrumental; integrative)	Mixed	Questionnaires (quant + qual); Interviews	1: What are the initial reasons or orientations of U.S. students for learning Arabic as a foreign language in a beginner-level course? 2: Will these orientations develop into sustained motivation to learn Arabic throughout the course of the semester? If so, do these sustained motivations fall in the instrumental/ integrative motivation dichotomy? Or are there other types, such as the social motivation? 3: Do the materials presented during the class help to sustain the learners' motivation? Why or why not? What are other materials and experiences both inside and outside the classroom that affect the learners' motivation?	Participants attracted to learn Arabic for personal development reasons, e.g., understanding a different culture, understand the world and travel to an Arab country; US students more communication oriented – have shown great interest in conversing with Arabic speakers and building friendships; Proposes expanding the classic integrative- instrumental dichotomy with addition of a social motivation component.
12	Husseinali	2012	Journal Article	Arabic Heritage Language Learners: motivation, expectations, competence, and engagement in learning Arabic	USA	University students (?)	50 (Arab HLs and Muslim HLs only)	Not specified	Quant	Questionnaire	Compare two categories of learners in terms of motivations, expected learning outcomes, perceived ability, and engagement.	HLs have strong identity and community motivations, but instrumental motivation is very weak; HLs more drawn to study Arabic to improve communication skills rather than to learn about heritage culture; Arabic HLs perceived themselves more competent to achieve speaking proficiency than Muslim HLs; Muslim HLs want to learn more about Islam/Islamic texts than Arab HLs.

13	Brosh	2013	Journal Article	Motivation of American College Students to Study Arabic	USA	College students	229	Gardner (?) (Integrative; Instrumental)	Quant	Questionnaire	Explores the motivations of American college students to study Arabic	The most commonly chosen reasons for studying: "Interest in the language", "Employment", "It is required for my major", and "Interest in the Arab culture/history". Five categories of motivations: integrative orientation; practical orientations based on a perception of the language's usefulness: economic utility and academic utility; psychological and social factors: national threat and ethnic heritage.
14	Alkhateeb	2014	Journal Article	Reading Anxiety, Classroom Anxiety, Language Motivation, Reader Self-Perception, and Arabic Achievement of Arab-American Students Learning Arabic as a Second Language	USA	Middle School Students (Muslim)	118	Gardner (AMTB – adapted)	Quant	Questionnaire; teacher-reported grades	Assessed the relations between reading anxiety, classroom anxiety, language motivation, and readers' self-perception	Significant positive correlations between language motivation and reader self-perception scores; Classroom anxiety, language motivation, and reader self-perception significantly predicted Arabic achievement; Learners expressed the importance of learning Arabic for understanding the Quran, and agreed that being a Muslim was a motivational factor to learn Arabic; The item "My parents have stressed the importance Arabic will have on understanding the Quran" (M = 4.05) significantly correlated with Arabic Achievement; The item "My parents feel that since I

												am a Muslim, I should learn Arabic" (M = 4.03) also significantly correlated with Arabic Achievement.
15	Meshkat & Saeb	2014	Journal Article	High-school Students' Beliefs about Learning English and Arabic	Iran	High school students (female)	540	Horowitz - Beliefs about Language Learning Inventory (BALLI) – includes motivation and expectations section	Quant	Questionnaire	Examined whether the target language has any effect on the students' beliefs about language learning	Learners' beliefs about difficulty of learning English and Arabic were not significantly different; Higher motivation to learn English.
16	Ajape et al.	2015	Journal Article	Students' Motivation and Attitude towards the Learning of Arabic Language: A Case Study of Arabic Students in Nigerian Universities	Nigeria	University students	288	Gardner's AMTB		Questionnaire	1) How do students perceive their attitude and motivation towards the learning of Arabic language in the Nigerian universities? 2) What are the relationships between the factors of attitude and motivation towards the learning of Arabic language?	High levels of integrativeness; low levels of instrumentality.
17	Ramezanzadeh	2015	MSc Diss	Exploring GCSE Arabic: An analysis of heritage and non- heritage language learner performance.	UK	Secondary school learners (year 10 – GCSE)	75	Gardner (Integrative; Instrumental)	Mixed	Questionnaires; Interviews; Proficiency Tests	Primary focus was a comparison of heritage and non-heritage performance and experience.	High levels of demotivation - almost half (46.8%) expressed a desire to drop the subject if possible; The main instrumental orientation, and the most prominent overall, was Religious; Career-related were almost negligible (4%); Integrative orientations also low. Almost 20% of students indicated that they were only studying Arabic because it was compulsory.

18	Semaan & Yamazaki	2015	Journal Article	The Relationship Between Global Competence and Language Learning Motivation: An Empirical Study in Critical Language Classrooms	USA	University students (of critical languages – including Arabic)	137	Gardner's AMTB	Quant	Questionnaire	'To determine the [...] relationship between (1) students' overall global competence and second language motivation, and (2) students' overall global competence and each of the subcategories of motivation (integrativeness, instrumental orientation, attitudes and beliefs, and parental encouragement).'	Positive relationship between global competence and L2 motivation, as well as integrativeness, instrumental orientation, and attitudes/beliefs.'
19	Zabarah	2015	Journal Article	College-Level Arabic Heritage Learners: Do they belong in Separate Classrooms?	USA	College heritage learners	57	Not specified (Integrative; instrumental)	Mixed (longitudinal)	Questionnaire (with open-ended Q for motivation)	examined the needs, reasons, and linguistic levels of Arabic heritage students at a major university.	Only one student reflected pure instrumental orientation; Live/ work/ study in region was most popular answer; Religion relatively low.
20	Djafri & Wimbari	2018	Journal Article	Measuring foreign language anxiety among learners of different foreign languages: in relation to motivation and perception of teacher's behaviors	Indonesia	University students	182	Noels et al. (2003) - Language Learning Orientation Scale (LLOS)	Quant	Questionnaire	1) To what extent does foreign language anxiety is manifested in Indonesian learners of English, Japanese, Korean, French, and Arabic? 2) Is there a difference in foreign language anxiety level between Latin languages group (English and French) and non-Latin language group (Japanese, Korean, and Arabic)? 3) How did the foreign language anxiety relate to students' motivation and their perception of teachers' behaviors among Indonesian learners of different foreign languages?	Learners of non-Latin languages group have higher foreign language anxiety than Latin language group; Arabic learners exhibit relatively high levels of foreign language anxiety compared to French, Korean and English, despite the fact that Arabic has been a familiar foreign language learned by most Indonesian people since childhood.

21	Abdelhalim & Alqubayshi	2020	Journal Article	Motivational Orientation and Language Acculturation Experienced by English Speaking Adults Learning Arabic in Saudi Arabia	Saudi Arabia	Students enrolled in Arabic course at a university (not clear whether they are university students)	50	Mix - The Motivation and Attitude Questionnaire developed by Dörnyei (1990); the Language Learning Orientations Scale (LLOS; Noels et al., 2000); Midraj et al., (2015); Öztürk & Gürbüz , 2013)	Mixed	Questionnaire; Interviews	1. What is the level and type of participants' motivational orientation (integrative – instrumental) towards learning ASL? 2. What kind of strategies can be adopted by ASL learners in the process of socialization and acculturation in Saudi society? 3. What is the level of language acculturative stress perceived by the participants? 4. In what ways do participants' socio-demographic data (cultural background, age, & length of residence in the hosting society), Arabic language proficiency level, motivational orientation toward learning it, and perceived acculturative stress correlate to their language acculturation strategy in the hosting society?	Integrative orientation to learning Arabic higher than instrumental motivation; Participants' motivational orientation level towards learning Arabic generally moderate; 55% reported that their main aim is to be able to communicate in Arabic and understand Saudi people; Second reason (36%) was to learn the language of the Quran.
L2MSS Studies involving Arabic Learners												
22	Al Khalil	2011	PhD Thesis	Second Language Motivation: Its Relationship to Noticing, Affect, and Production in Task-based Interaction	USA	University students (?)	44	Combination: Gardner's S-EM; Situated perspective of state motivation; L2MSS	Mixed	Questionnaire; video-recorded task-based interaction; stimulated recall	What are the relationships between L2 motivation and noticing, task affect, and L2 production in oral task- based interaction?	Attitudes towards the L2 community significantly predicted reported noticing of recasts; State motivation significantly correlated with task satisfaction and task relevance; Motivation, as defined by S-EM, significantly predicted accuracy, complexity, and fluency in four

												participants, the two most-motivated and the two least-motivated in the sample; No measures from the L2MSS associated with noticing of recasts or L2 production.
23	Anya	2011	Journal Article	Connecting with communities of learners and speakers: Integrative ideals, experiences, and motivations of successful black second language learners.	USA	African-American college students	6 (1 Arabic)	L2MSS	Qual	Language learning autobiographies	1. How do successful black college L2 learners describe their home background and experience with languages, and do they express any connection between those and their choices/efforts in SLA? 2. How do the learners describe both their naturalistic and formal classroom experiences in SLA? 3. What reasons do the learners provide for (not) wanting to study foreign languages in college? 4. How do the learners address issues of race or ethnicity as related to SLA?	
24	Thompson	2017	Journal Article	Don't tell me what to do! The anti-ought-to self and language learning motivation	USA	L1-English speaking advanced learners of Arabic and Chinese	2	L2MSS; Complex Dynamic Systems Theory (CDST)	Qual	Interviews (Narrative Enquiry)	'explore the relationship of the L2 Motivational Self System (L2MSS) and Complex Dynamic Systems Theory (CDST), specifically examining the role of context in developing the different selves of language learners.'	Anti-ought-to self has a "dominant" relationship with the context, in contrast to the "submissive" relationship of the ought-to self; 'Dynamicity' of the psychological aspects of self - selves fluctuate throughout the language learning experience.
25	Kong et al.	2018	Journal Article	L2 Motivational Self System, international	South Korea	College students	1296 (658 studyin	L2MSS	Quant	Questionnaire	1. Is 'the L2 Motivational Self System' a valid tool for understanding and explaining the	L2 learning attitude most influential factor on learners' intended effort, followed by the IS;

				posture and competitiveness of Korean CTL and LCTL college learners: A structural equation modeling approach			g LCTLs; 401 Arabic)				L2 motivation of Korean CTL and LCTL learners? a) Which of the three motivational components of the L2 Motivational Self System best predicts the efforts of CTL and LCTL learners? b) Does the ideal L2 self better explain the efforts of CTL and LCTL learners than the ought-to L2 self? 2. How does the impact of international posture and competitiveness on the three motivational components of the L2 Motivational Self System differ between CTL and LCTL groups?	Impact of the ought-to L2 self on L2 learning effort was relatively low.
26	Blair & Azaz	2019	Journal Article	Self-guide distinctions and episodic memory in L2 motivation: The case of Arabic in the United States	USA	University students	54	L2MSS (with integration of episodic memory)	Quant	Questionnaire	1. How does ought-to L2 self/own and ought-to L2 self/others function within the L2 Motivational Self System among L2 Arabic learners in the United States? 2. To what extent is episodic memory associated with L2 self-guides?	L2LE had strongest direct effect and IS had strongest indirect effect on intended effort; L2 self-guides most influential when mediated by L2LE or IS; OTS/own OTS/others exhibited dichotomous behaviour - OTS/own had a positive effect and ought-to L2 self/others had a negative effect on intended effort; Associations between episodic memory and self-guides were found - but these were selective, possibly driven by shared emotional and self-relevant content.

27	Calafato & Tang	2019	Journal Article	The status of Arabic, superdiversity, and language learning motivation among non-Arab expats in the Gulf	UAE	Secondary school students (non-Arab expatriates attending private UK curriculum schools)	114	L2MSS	Quant	Questionnaire	How does motivation to learn Arabic compare with motivation to learn English? 1.1 How does it compare with motivation to learn an additional foreign language? Does English produce negative interference with regard to motivation to learn LOTE's? In terms of the desire to be multilingual, what motivating factors differentiate learners studying three languages simultaneously from those studying only two?	Participants approach Arabic study in mostly functional terms, not particularly interested in Arabic media or culture. Parental influence does not appear to be important motivating factor; Participants' motivational selves hint at lukewarm attitude - do not really see a strong reason to study it; Participants' IS displays the least motivational intensity of all three languages; Participants express strong levels of motivation to learn Arabic for education/training purposes (a compulsory subject at school); They also consider Arabic to be important for their careers, which indicates that they might be planning to find work in the UAE.
28	Calafato	2020	Journal Article	Learning Arabic in Scandinavia: Motivation, metacognition, and autonomy	Scandinavia (Norway, Sweden, Denmark)	University students	96	Partly L2MSS	Quant	Questionnaire (inc. open-ended questions)	1. What factors motivate learners in Norway, Sweden, and Denmark to learn Arabic? 1.a. Are there any differences between the participants based on context? 2. To what extent do they self-regulate their learning of Arabic?	Participants differed statistically significantly in motivation to learn Arabic; Participants do not have distinct motivational selves, either IS or OTS – proposes a domain-specific conceptualization of motivation without the inclusion of selves (Nakamura, 2019).

B – METHODOLOGY

B.1. Population of Schools teaching GCSE Arabic

Based on the data available, there are currently 144 schools offering Arabic in England. Of these 112 (77.8%) were Muslim faith schools,⁴³ with the remaining 32 (22.2%) a mixture of other faith and non-denominational schools.

For each school, information was collected on the following: gender of students; region; school type; religious affiliation; Ofsted rating. Data for the entire population is presented for each of these categories in Table B.1.1., followed by a breakdown and comparison of Muslim faith schools and Other schools.

Table B.1.1. Population of Schools teaching GCSE Arabic						
	Muslim	%	Other	%	Total	%
<i>Gender</i>						
Girls	61	54.5	12	37.5	73	50.7
Boys	40	35.7	3	9.4	43	29.9
Co-Ed	11	9.8	17	53.1	28	19.4
<i>School Type</i>						
Independent	96	85.7	11	34.4	107	74.3
Voluntary Aided (VA)	4	3.6	0	0.0	4	2.8
Academy	12	10.7	11	34.4	23	16.0
Community	0	0.0	6	18.8	6	4.2
Foundation	0	0.0	4	12.5	4	2.8
<i>Ofsted Rating</i>						
Outstanding	22	19.6	9	28.1	31	21.5
Good	49	43.8	13	40.6	62	43.1
Requires Improvement	21	18.8	1	3.1	22	15.3
Inadequate	17	15.2	0	0.0	17	11.8
No report	3	2.7	9	28.1	12	8.3
<i>Region</i>						
North East	1	0.9	1	3.1	2	1.4
North West	29	25.9	2	6.3	31	21.5
Yorkshire and the Humber	17	15.2	3	9.4	20	13.9
East Midlands	9	8.0	0	0.0	9	6.3
West Midlands	21	18.8	4	12.5	25	17.4
East Anglia	4	3.6	2	6.3	6	4.2
London	27	24.1	11	34.4	38	26.4
South East	2	1.8	6	18.8	8	5.6
South West	2	1.8	3	9.4	5	3.5

⁴³ This information was correct at the time of collation, which was conducted between 2016-17.

Gender

Just over half of all schools offering Arabic are single-sex girls schools (50.7%), followed by single sex boys schools (29.9%). Only 1 in 5 schools offering the GCSE as a taught option were co-ed (see 'Total column' in Table B.1.2).

Table B.1.2. Population of Schools teaching GCSE Arabic – Gender Breakdown						
	Muslim	%	Other	%	Total	%
Girls	61	54.5	12	37.5	73	50.7
Boys	40	35.7	3	9.4	43	29.9
Co-Ed	11	9.8	17	53.1	28	19.4

Muslim Schools

The frequency of girls schools in the Muslim school population clearly contributes to the overall bias towards this type of school in the total. Most Muslim schools are single sex, and the majority of these are girls schools. Co-ed Muslim schools are relatively rare, with only 11 of 144 educating both boys and girls in the same institution. Incidentally, this creates an equity issue – girls are more likely to learn Arabic because there are more Islamic girls' schools.

Other Schools

The majority of non-Muslim schools offering Arabic are co-educational (17 out of 32), as reflects the figures available from the Gov.uk database on educational institutions for the population of such schools in England. However, as many of the schools in this specific population are private institutions, amongst which single sex education is more common, the proportion of single sex girls' schools within the database is higher than the England average.

School Type

The vast majority of the population of schools are independent (see Table B.1.3). This is explained by a number of factors, including the fact that independent schools have more freedom in determining their curriculum, and so there may be less restrictions in offering a 'Less Commonly Taught Language'.

Table B.1.3. Population of Schools teaching GCSE Arabic – School Type Breakdown						
	Muslim	%	Other	%	Total	%
Independent	96	85.7	11	34.4	107	74.3
VA	4	3.6	0	0.0	4	2.8
Academy	12	10.7	11	34.4	23	16.0
Community	0	0.0	6	18.8	6	4.2
Foundation	0	0.0	4	12.5	4	2.8

Muslim Schools

The majority of Islamic schools are independent since, as aforementioned, this allows the governing body the freedom to integrate Islamic subjects into the curriculum, and deliver education within an Islamic ethos (defined individually by each school). Whilst the label ‘independent school’ is often associated with selective, high-fee charging institutions, Islamic schools tend to charge minimal annual fees. For the schools in this population, the range of annual fees varies from £850 to £6,900, with an average fee of £2,445 per student (based on the lower band of fees for each school)⁴⁴. Of those schools who fell into the VA or Academy categories, several of these had transitioned from being independent, after applying for Academy/ VA status.

Other Schools

Of the remaining schools, just over a third were independent, with the remainder falling under the category of state funded schools. None of these schools were Voluntary Aided. The range of annual fees for the independent schools was in line with the general profile of private establishments, with yearly tuition fees ranging between £12,532 and £42,501, with an average of £24,234.

Ofsted Rating

The most recent Ofsted report for each school was examined where available, and the overall rating recorded. The majority of schools in the population (64.6%) were rated as either ‘good’ or ‘outstanding’ (see Table B.1.4). Over a quarter, however, were considered by Ofsted to be ‘inadequate’ or ‘requiring improvement’. Since independent schools are not subject to Ofsted inspections, ratings for 12 of these schools are absent.

Table B.1.4. Population of Schools teaching GCSE Arabic – Ofsted Rating Breakdown						
	Muslim	%	Other	%	Total	%
Outstanding	22	19.6	9	28.1	31	21.5
Good	49	43.8	13	40.6	62	43.1
Requires Improvement	21	18.8	1	3.1	22	15.3
Inadequate	17	15.2	0	0.0	17	11.8
No report	3	2.7	9	28.1	12	8.3

⁴⁴ Figures generated from a survey of these school websites. The lower band of fees generally represents the annual fee for tuition per student, whereas higher bands include boarding costs.

Muslim Schools

Though the majority of Muslim schools in this population received a 'good' or 'outstanding' rating in their last inspection, a worrying 34% of schools were deemed not to meet the expected minimum standards or required improvement. Often Muslim schools are set up by local faith groups or individuals and serve lower income communities. This often severely limits their access to funds, and thus restricts investment in high quality resources, personnel, and educational experiences and opportunities for students. Independent schools are exempt from the requirement to only employ qualified teachers, which may also have an impact on standards.

Other Schools

Only one of the schools in the Other category was rated as requiring improvement. Of all remaining 22 schools who underwent an Ofsted inspection, none of them were considered inadequate.

Region

Just over a quarter of all schools offering Arabic were situated in the London region (see Table B.1.5). This is expected both in terms of the population of the region, which has a high proportion of Muslim and Arab residents in a relatively concentrated and accessible area, both making it more likely for Islamic schools to be established there, and for Other schools to have higher populations of Muslim/ Arab students making offering the subject a more viable option. The same could be said for the North West, which contains cities such as Manchester and Lancashire which have high Muslim and Arab populations.

Table B.1.5. Population of Schools teaching GCSE Arabic – Regional Breakdown						
	Muslim	%	Other	%	Total	%
North East	1	0.9	1	3.1	2	1.4
North West	29	25.9	2	6.3	31	21.5
Yorkshire	17	15.2	3	9.4	20	13.9
East Midlands	9	8.0	0	0.0	9	6.3
West Midlands	21	18.8	4	12.5	25	17.4
East Anglia	4	3.6	2	6.3	6	4.2
London	27	24.1	11	34.4	38	26.4
South East	2	1.8	6	18.8	8	5.6
South West	2	1.8	3	9.4	5	3.5

Muslim Schools

The majority of Muslim schools, 29 out of 112, are located in the North West. This is followed closely by London with 27 schools. There are very few Muslim schools in the

south of England, which again reflects the national concentrations of Muslim populations.

Other Schools

Amongst Other schools, most, 11 out of 32, are situated in London. Interestingly there are more schools in the South of England offering the subject. In contrast to the proportion of Muslim schools in the North West, there is a far lower percentage of Other schools, with only 2 out of 32 schools situated in Manchester and no other city in that region.

DEPARTMENT OF EDUCATION



Arabic Language Survey

Please fill in the details below. Don't worry, this slip will be cut off and thrown away once the data has been digitised. Your information will be kept completely anonymous and confidential!

Name: _____

School: _____ Year Group: _____

Date of Birth: _____ Gender: Female ☐ Male ☐

Survey Instructions

Thank you for agreeing to take part in my study! Please take a few minutes to read these instructions carefully before you start!

To start, please make sure you have completed the details on the front cover.

Then, please complete the questions inside this booklet to the best of your ability. If a question is in *italics*, it means you only have to answer it if you ticked 'yes' (or a specified answer) on the question before. If you have any problems, feel free to ask the researcher or your teacher to explain.

If you don't feel comfortable answering any of the questions you can leave them blank, though the more responses you can provide, the more helpful it will be!

Feel free to jot down any notes or extra information about any of your answers that you think might be useful for the researcher.

Remember, you are free to withdraw from the study at any point with no negative consequences. All your answers will be kept anonymous and confidential, and will not affect your school grades in any way.

Thanks again!

1) How many years have you studied Arabic **at school** in total (at your current school or previous ones) **not** including the current year?

Less than 1	1	2	3	4	More than 5
☐	☐	☐	☐	☐	☐

2) Have you ever studied Arabic **outside** of school?

☐ Yes ☐ No

2a) If you answered 'yes' to the previous question, where did you study Arabic outside of school? (Tick as many as necessary.)

☐ At the mosque	☐ Studied the 'Alimah course
☐ At home with parents	☐ At an afterschool / evening / weekend class
☐ At home with a tutor	
☐ Other (Please specify where: _____)	

3) How do you feel about the Arabic language? Tick one circle on the scale below:


☐
☐
☐
☐
☐
☐
☐


4) How do you feel about studying the Arabic GCSE?


☐
☐
☐
☐
☐
☐
☐


5) Why are you studying the Arabic GCSE?

6) What is your ethnicity?

White or White British

- ☐ White English/ Scottish/ Irish
☐ Any other white background

Asian or Asian British

- ☐ Pakistani
☐ Bangladeshi
☐ Indian
☐ Any other Asian background

Arab or Arab British

- ☐ Arab/ Arab British (see q. 6a)

Black or Black British

- ☐ African
☐ Caribbean
☐ Any other black background

Mixed or Mixed British (see q. 6b)

- ☐ White and Asian
☐ White and Arab
☐ White and Black African
☐ White and Black Caribbean
☐ Any other mixed background

6a) If you ticked 'Arab or Arab British' please indicate your country of origin:

6b) If you ticked 'Mixed or Mixed British' please indicate where your parents are/were from:

7) On the scales below, tick the circle that is most true of you:

I feel strongly connected to my ethnicity	☐ ☐ ☐ ☐ ☐ ☐	I don't feel connected to my ethnicity at all
I know a lot about my culture and heritage	☐ ☐ ☐ ☐ ☐ ☐	I don't know anything about my culture and heritage
My ethnicity plays a big role in how I see myself	☐ ☐ ☐ ☐ ☐ ☐	My ethnicity doesn't play a role in how I see myself at all

8) What languages are spoken in your home?

9) Think about the following statement, tick the circle in each line that is most true of you:

In the future, I will be someone who can fully understand religious texts in Arabic.

There will be bad consequences if I don't achieve this	☐	☐	☐	☐	☐	There won't be bad consequences if I don't achieve this
This is very likely	☐	☐	☐	☐	☐	This is very unlikely
I can fully understand religious texts in Arabic	☐	☐	☐	☐	☐	I can't understand religious texts in Arabic at all
I imagine this very often	☐	☐	☐	☐	☐	I never imagine this
I desire this very much	☐	☐	☐	☐	☐	I do not desire this at all
This is true of me now	☐	☐	☐	☐	☐	This is not true of me at all right now
My imaginations about this are very clear and real to me	☐	☐	☐	☐	☐	My imaginations about this are very vague
People around me will be disappointed if I don't achieve this	☐	☐	☐	☐	☐	People around me will be fine if I don't achieve this
I feel forced to achieve this	☐	☐	☐	☐	☐	I don't feel forced to achieve this
Other people want me to achieve this goal	☐	☐	☐	☐	☐	No one else wants me to achieve this goal
I am very confident in my ability to understand religious texts in Arabic	☐	☐	☐	☐	☐	I am not confident in my ability to understand religious texts in Arabic

10) Do you speak any Arabic at home?

☐ Yes ☐ No

11) Is Arabic spoken in your home?

☐ Yes ☐ Not currently, but it used to be ☐ No

11a) If you answered 'Yes' or 'Not currently', to question 11, please indicate who speaks/spoke Arabic in your home:

11b) If you answered 'Yes' to question 10 or 11, what type of Arabic is used?

☐ Standard Arabic (*Fuṣṣḥa*) ☐ An Arabic dialect ('*Ammiyyah*)
☐ Both ☐ I don't know

12) Tick which one of the following apply to you:

☐ Both my parents speak/spoke Arabic.
☐ One of my parents speaks/spoke Arabic.
☐ Only my grandparents speak/spoke Arabic.
☐ None of the above.

13) Have you ever lived in an Arab country (for longer than 3 months at a time)?

☐ Yes ☐ No

13a) If you answered 'Yes', how old were you when you lived there?

14) Do you have relatives who live in Arab countries?

☐ Yes ☐ No

14a) If you answered 'Yes', how often do you visit them?

15) Think about the following statement, then tick the circle in each line that is most true of you:

In the future, I will be someone who can communicate fluently with Arabic speakers.

I can communicate fluently with Arabic speakers	☐	☐	☐	☐	☐	☐	I can't communicate with Arabic speakers at all
I feel forced to achieve this	☐	☐	☐	☐	☐	☐	I don't feel forced to achieve this
This is true of me now	☐	☐	☐	☐	☐	☐	This is not true of me at all right now
My imaginations about this are very clear and real to me	☐	☐	☐	☐	☐	☐	My imaginations about this are very vague
I imagine this very often	☐	☐	☐	☐	☐	☐	I never imagine this
People around me will be disappointed if I don't achieve this	☐	☐	☐	☐	☐	☐	People around me will be fine if I don't achieve this
This is very likely	☐	☐	☐	☐	☐	☐	This is very unlikely
Other people want me to achieve this goal	☐	☐	☐	☐	☐	☐	No one else wants me to achieve this goal
I am very confident in my ability to communicate with Arabic speakers	☐	☐	☐	☐	☐	☐	I am not confident in my ability to communicate with Arabic speakers
I desire this very much	☐	☐	☐	☐	☐	☐	I do not desire this at all
There will be bad consequences if I don't achieve this	☐	☐	☐	☐	☐	☐	There won't be bad consequences if I don't achieve this

16) Do you have any Arabic speakers in the community where you live?

☐ Yes ☐ No ☐ I don't know

17) Do you have any contact with Arabic speakers in your community?

☐ Yes ☐ No

18) Do you use internet, mobiles, or social media to contact friends/ family in Arabic?

☐ Yes ☐ No

19) Which religion do you follow?

☐ Islam	☐ Buddhism
☐ Christianity	☐ Sikhism
☐ Judaism	☐ None
☐ Hinduism	☐ Other (Please specify: _____)

19a) If you follow a religion, on the scales below, tick the circle that is most true of you:

I feel strongly connected to my religion	☐ ☐ ☐ ☐ ☐ ☐	I don't feel connected to my religion at all
I consider myself a very practicing religious person	☐ ☐ ☐ ☐ ☐ ☐	I don't consider myself a very practicing religious person at all
Religion is very important to my identity	☐ ☐ ☐ ☐ ☐ ☐	Religion is not important to my identity at all

20) In your own words, describe how you feel about studying the Arabic GCSE?

Note: **Standard Arabic** refers to *Modern Standard Arabic* or **fusha** (as used in the news), **Arabic dialects** are known as **ammiyyah**, which are often spoken at home and differ from country to country.

21) For each line below tick the square on the scale that represents your level right now:

I can read Arabic words out loud correctly.	☐	☐	☐	☐	☐	☐	I can't read Arabic words at all.
I can write short texts in standard Arabic on a range of familiar topics.	☐	☐	☐	☐	☐	☐	I can't write anything in standard Arabic .
I can speak in standard Arabic to talk about a range of familiar topics.	☐	☐	☐	☐	☐	☐	I can't speak any standard Arabic .
I can understand written texts that talk about familiar topics in standard Arabic .	☐	☐	☐	☐	☐	☐	I can't understand anything written in standard Arabic .
I can understand the main points of clear standard speech in Arabic on familiar matters.	☐	☐	☐	☐	☐	☐	I can't understand anything I hear in standard Arabic .
I can tell the difference between standard Arabic and an Arabic dialect .	☐	☐	☐	☐	☐	☐	I can't tell the difference between standard Arabic and Arabic dialects at all.
I can speak in an Arabic dialect about a range of familiar topics.	☐	☐	☐	☐	☐	☐	I can't speak in an Arabic dialect at all.
I can understand the main points when I hear someone speaking in an Arabic dialect about familiar matters.	☐	☐	☐	☐	☐	☐	I can't understand anything I hear in an Arabic dialect .
When I hear the Quran read out loud, I can understand the main topic.	☐	☐	☐	☐	☐	☐	When I hear the Quran read out loud, I can't understand anything.
I can read Arabic words when they have harakat on them.	☐	☐	☐	☐	☐	☐	I can't read anything in Arabic at all.
When I read the Quran in Arabic, I can understand the main topic of each verse.	☐	☐	☐	☐	☐	☐	I can't understand anything in the Quran in Arabic.

22) How does your Arabic grade compare to your other GCSE subjects?

- ☐ I am doing better in Arabic than my other subjects
- ☐ I'm doing worse in Arabic than my other subjects
- ☐ I am doing the same in Arabic as I am in my other subjects
- ☐ I don't know how my Arabic grade compares to my other subjects.
- ☐ Other (Please give details: _____)

23) Do you see yourself using the Arabic language in your life in the future?

☐ Yes ☐ No

23a) If you answered 'Yes' to question 23, please say what you see yourself using Arabic for:

23b) Has the GCSE helped you to do these things so far?



24) Would you still study the Arabic GCSE if you had the choice?

☐ Yes ☐ No

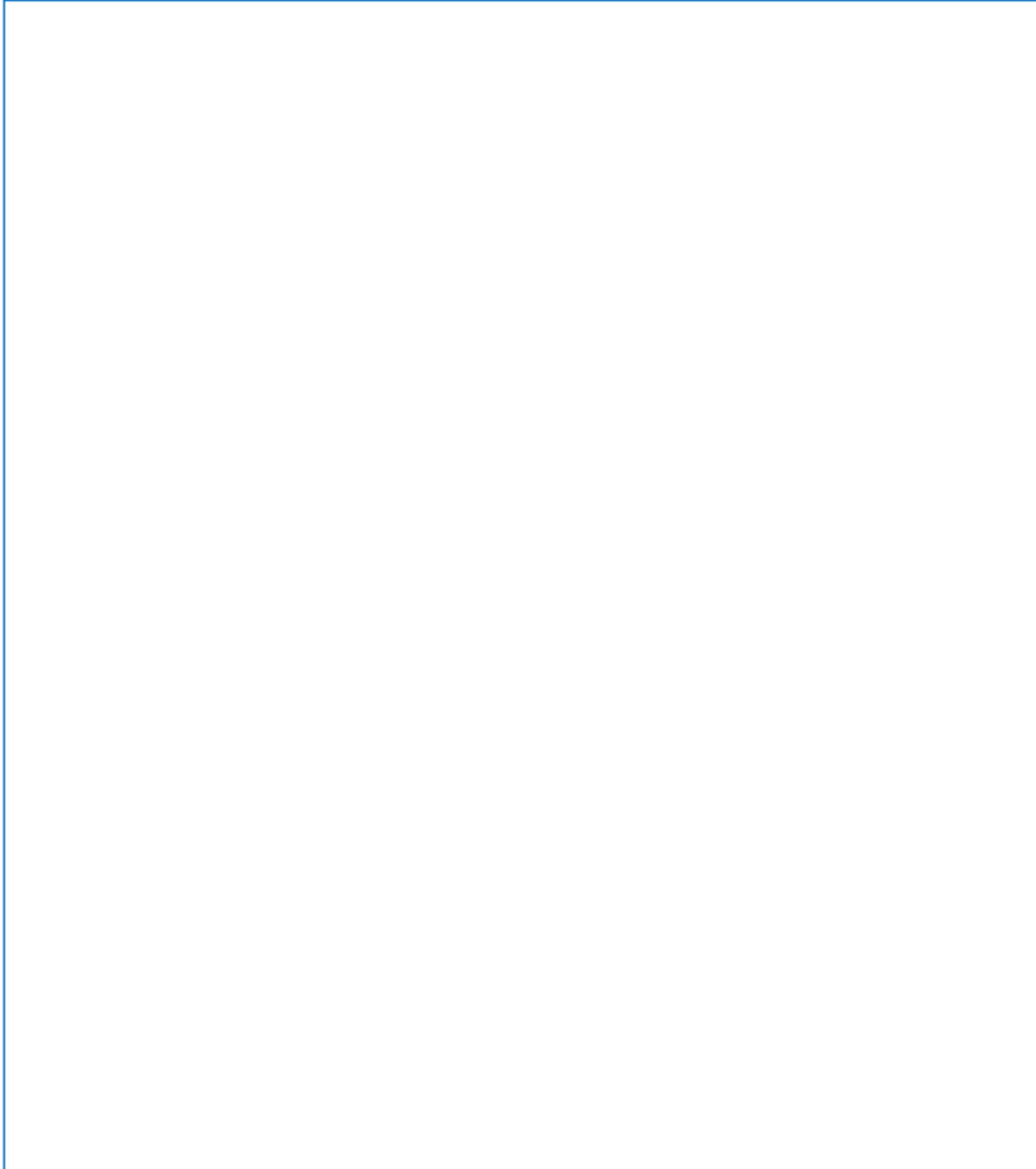
25) Why/ why not?

26) Think about the following statement, then tick the circle in each line that is most true of you:

In the future, I will be someone who can use Arabic to fully understand the news, books, or magazines.

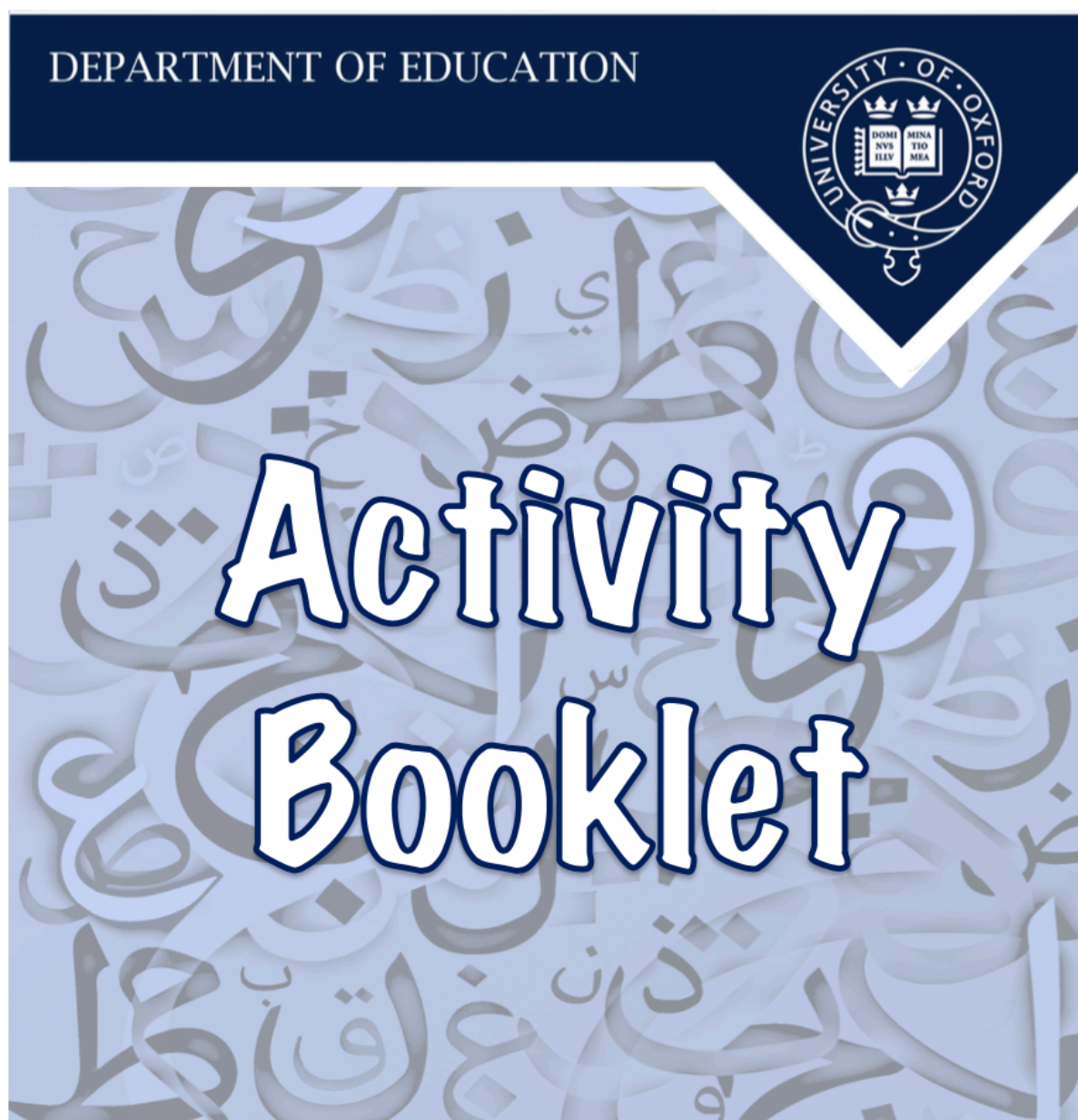
My imaginations about this are very clear and real to me	☐	☐	☐	☐	☐	My imaginations about this are very vague
Other people want me to achieve this goal	☐	☐	☐	☐	☐	No one else wants me to achieve this goal
This is true of me now	☐	☐	☐	☐	☐	This is not true of me at all right now
There will be bad consequences if I don't achieve this	☐	☐	☐	☐	☐	There won't be bad consequences if I don't achieve this
I imagine this very often	☐	☐	☐	☐	☐	I never imagine this
People around me will be disappointed if I don't achieve this	☐	☐	☐	☐	☐	People around me will be fine if I don't achieve this
This is very likely	☐	☐	☐	☐	☐	This is very unlikely
I am very confident in my ability to understand Arabic news/ books/ magazines	☐	☐	☐	☐	☐	I am not confident in my ability to understand Arabic news/ books/ magazines
I feel forced to achieve this	☐	☐	☐	☐	☐	I don't feel forced to achieve this
I desire this very much	☐	☐	☐	☐	☐	I do not desire this at all
I can fully understand the news/ books/ magazines in Arabic	☐	☐	☐	☐	☐	I don't understand the news/ books/ magazines in Arabic at all

27) Do you have any other comments about learning Arabic, or the GCSE? For example, this could be related to your experiences learning Arabic, what you think of the materials or exams, the resources you use, what you'd like more/ less of in the classroom, or anything else!



***Thank you very much for taking the time to complete this survey!
Your answers are very important to my study.***

B.3. Proficiency Tests



Please fill in the details below. Don't worry, this slip will be cut off and thrown away once the data has been digitised. Your information and answers will be kept anonymous and confidential.

Name: _____

School: _____ Year Group: _____

Date of Birth: _____ Gender: Female ☐ Male ☐

Activity 1: Do you know this word?

Instructions

On the next pages is a list of words in Arabic.

If you **know** the word, **tick** the box next to it.

If you **don't know** the word, or are **not sure**, leave the box **blank**.

Here's an example:

☒	ليمون
☐	موز
☐	كرز

This student knows the word ليمون, so she ticked the box. She's not sure if she knows the word موز, so she left the box blank. She doesn't know the word كرز, so she left that box blank too.

Please remember that this is **individual** work, so please don't ask your friends or your teacher for help with the answers!

If you have any questions, feel free to ask the researcher!

And don't worry, your answers to these will be kept anonymous and will not affect your school grades in any way!

Activity 2: Fill-in-the-Gap Tasks

Instructions

On the following pages are some sentences in Arabic.

Your job is to pick the underlined word that fits each sentence.
Please circle your choice.

Here's an example:

ذهبنا أنا وأصدقائي إلى المسجد \ المطعم \ المكتبة لنأكل الغداء.

In this sentence, the word المطعم completes the sentence correctly. The other two words don't make sense in this gap.

Please try to have a go at them all.

Please remember that this is **individual** work, so please don't ask your friends or your teacher for help with the answers!

If you have any questions, feel free to ask the researcher!

And don't worry, your answers to these will be kept anonymous and will not affect your school grades in any way!

Section 1

(2017 Sample - FT Q10)

صَبَّاحَ كُلِّ يَوْمٍ، أُحِبُّ أَنْ أَتَنَاوَلَ بَيْتاً \ فُطُوراً \ عَشَاءً خَفِيفاً جِداً. أَكُلُّ
الْبَيْضِ وَالشَّايِ مَعَ الْحَلِيبِ \ الزَّيْتِ \ التُّفَاحِ.

تَتَنَاوَلُ عَائِلَتِي الْغَدَاءَ فِي السَّاعَةِ الْوَاحِدَةِ، لِأَنَّنِي \ هَذِهِ \ وَلَكِنِّي
لَا أَكُونُ مَعَهُمْ. أَتَنَاوَلُ غَدَاءً سَرِيعاً إِلَى \ مِنْ \ فِي مَكْتَبِي.

أُحِبُّ أَنْ أَتَنَاوَلَ الْعَشَاءَ مَعَ وَجِبَتِي \ غُرْفَتِي \ عَائِلَتِي فِي وَقْتٍ مُبَكِّرٍ كُلَّ
مَسَاءٍ. أَتَنَاوَلُ الْعَشَاءَ \ الْغَدَاءَ \ الْمَدْرَسَةَ فِي السَّاعَةِ السَّابِعَةِ كُلَّ يَوْمٍ.

وَجِبَتِي الْمُفَضَّلَةُ \ الْمُحِبَّةُ \ الْجَمِيلَةُ هِيَ وَجِبَةُ الْغَدَاءِ. أَتَنَاوَلُ غَدَائِي كُلَّ
سَاعَةٍ \ يَوْمٍ \ دَقِيقَةً فِي مَطْعَمِ الْمَدْرَسَةِ.

Section 2

(2017 Sample - FT Q10)

في وقت فراغي أحب أن ألعب طعام \ غرفة \ كرة السلة. أذهب إلى
النادي الرياضي القريب في على \ من بيتنا بالحافلة رقم تسعة.
تستغرق الرحلة حوالي \ خلف \ بجانب ربع ساعة. السبت الماضي،
ذهبت مع كتابي \ أخي \ مدرّسي الصغير للسباحة. الرياضة مفيدة
لأنها تقوّي البلد \ الجسم \ اللعبة.

Section 3

(2017 Sample - HT Q3)

كنّا نسافر في إجازة الصيف كل معظم \ بعد عام إلى اليمن لزيارة
جدي وجدتي. نجلس \ نسافر \ نلعب بالطائرة والرحلة طويلة ومتعبة
جداً. نستمتع بالسباحة \ بالقراءة \ بالتنس في الطائرة أثناء السفر
ونفضّل أن نشاهد \ نأكل \ نقرأ أفلاماً بوليسية.

يعيش بعض أفراد عائلتي في شرق عطلة \ فرنسا \ مدرسة. نحن
نذهب لزيارتهم كل عطلة. نذهب بالطائرة \ بالقطار \ بالحاسوب لأن
أمي لا يُعجبها السفر بالطائرة. نقضي معظم \ قبل \ كثير الرحلة
باللعب على الألعاب الإلكترونية وبالتحدّث إلى \ في \ مع بعضنا، حيث
نخطّط اجازتنا في فرنسا.

يبعد بيت عمّتي حوالي ثلاث ساعات كل \ فقط \ بعض ولهذا نحن
نسافر لزيارتهم كل شهر. عادةً \ شخصياً \ مهماً نسافر بالقطار وأنا
أفضله لأنني أتعرف خلفه \ فيه \ فوقه على أصدقاء جدد، ولكننا
سوف نسافر بالبلد \ بالزيارة \ بالسيارة مع أبي في العطلة المقبلة.

You're all finished! Well done!

You can use any remaining time to check over your answers and then
hand your booklet back to the researcher or your teacher.

Thank you for taking part!



B.4. Proficiency Test Reliability Analysis

Y/N Items

The Y/N tests overall displayed a good level of internal consistency ($\alpha_{T1} = .97$, $\alpha_{T2} = .96$), as did the individual frequency bands whose alpha scores are represented in Table B.4.1, with scores ranging from $\alpha = .82$ to $.97$ at T1, and $\alpha = .80$ to $.96$ at T2. None of the resulting values for deleting individual items were sufficiently high enough to warrant removing them from the test, being only $.001 - .005$ higher than the original scale score (Field, 2018).

Time Point	Items	Mean	Variance	SD	N	N of Items	Cronbach's Alpha	Cronbach's Alpha based on Standardized Items	Suggested items to delete	Scale Mean if Item Deleted	Cronbach's Alpha if Item Deleted
T1	Y/N Test	24.08	239.347	15.471	264	60	.966	.966	-	-	-
	Level 0	7.55	6.948	2.636		10	.832	.837	-	-	-
	Level 1	4.34	10.567	3.251		11	.853	.855	-	-	-
	Level 2	2.87	8.399	2.898		10	.850	.852	-	-	-
	Level 3	3.07	8.436	2.905		10	.845	.848	5h	2.49	0.846
	Level 4	2.82	8.294	2.880		10	.851	.856	-	-	-
	Level 5	3.44	7.205	2.684		9	.818	.821	-	-	-
T2	Y/N Test	25.1	232.158	15.237	172	60	.963	.963	-	-	-
	Level 0	7.61	7.830	2.798		10	.864	.872	-	-	-
	Level 1	5.01	10.661	3.265		11	.838	.838	-	-	-
	Level 2	2.84	8.086	2.844		10	.840	.842	-	-	-
	Level 3	3.65	8.053	2.838		10	.816	.817	-	-	-
	Level 4	2.92	7.958	2.821		10	.834	.842	4d 5d	2.49 2.51	0.839 0.836
	Level 5	3.48	6.859	2.619		9	.802	.804	-	-	-

Cloze Tests

Each set of items had a good level of internal consistency, as determined by the Cronbach's alpha scores represented in Table B.4.2, as did the cloze tests overall at both time points ($\alpha_{T1} = .966$; $\alpha_{T2} = .918$). Item 3g again appears slightly problematic at both time points, with the alpha values suggesting its deletion should result in an increase in the overall validity of the test. However, since the improvements were only in the range of $.001 - .003$, these do not appear to warrant removing the items from the test without further investigation.

Table B.4.2. Reliability Scores per Cloze Test Activity T1&2											
Time Point	Items	Mean	Variance	SD	N	N of Items	Cronbach's Alpha	Cronbach's Alpha based on Standardised Items	Suggested items to delete	Scale Mean if Item Deleted	Cronbach's Alpha if Item Deleted
T1	Cloze Test	13.79	38.431	6.199	263	25	0.885	.885	3g	13.54	0.887
	Activity 1	5.13	4.823	2.196	263	8	0.72	0.721	1d	4.53	0.722
	Activity 2	2.86	2.157	1.469	263	5	0.609	0.613	-	-	-
	Activity 3	5.80	12.029	3.468	263	12	0.829	0.828	3g	5.55	0.831
T2	Cloze Test	15.43	45.846	6.771	171	25	0.918	.919	3g	15.13	0.919
	Activity 1	5.43	6.128	2.476	171	8	0.833	0.835	-	-	-
	Activity 2	3.22	2.268	1.506	171	5	0.668	0.676	-	-	-
	Activity 3	6.78	11.656	3.414	171	12	0.835	0.835	3g	6.48	0.838

B.5. Interview Protocol

Part 1: Greeting and General Guidelines

Thank you for agreeing to take part in this study and for sitting for this interview with me. I would like to remind you that the interview will be audio-recorded but all the data is confidential and will be kept anonymous. If there are any questions you don't understand or that you would like me to repeat, please let me know. Also, feel free to skip any questions you do not feel comfortable answering. If at any point you feel that you want to withdraw from this interview, you're completely free to do so.

Part 2: Questions

1. What has been your experience of studying the GCSE so far?
 - a. How confident/ comfortable are you with learning Arabic right now?
 - b. Would you say this has changed over the past year? Why?
2. Do you have a clear vision of who you want to be/ what you want to do in the future? What does this look like?
3. Do you see yourself using Arabic in the future? How?
 - a. Do you think about this much?
 - b. How clear is this vision?
 - c. How likely do you think it is that you'll achieve this?
 - d. Would you say this has changed over the past year? Why?
4. Do you feel pressured in any way to learn Arabic?
 - a. Is there anyone who thinks you should learn Arabic/ makes you feel like you should learn Arabic? Who is this, and why?
 - b. What would happen if you didn't become someone who could communicate/ understand Arabic?
 - c. Would you say this has changed over the past year? Why?
5. What role does religion play in your learning of Arabic, if any?
6. Is there anything in your Arabic classes that affects how well you're doing in Arabic?
7. Do you have any other comments or reflections on studying the Arabic GCSE?

Part 3: Thanks

Thank you for participating in this interview, your responses have been very helpful!

B.6. Procedure for Categorising Heritage Language Learners

As outlined in B.4.1 At T1, 26 students identified as being of Arab ethnicity. 75 participants responded 'yes' in answer to the question 'Do you speak Arabic at home?'. 49 identified Arabic as one of the languages spoken in their home, an open-ended question. 64 students indicated that Arabic was spoken in their home, with a further 17 saying that it used to be.

Table B.4.1 Arabic Heritage Indicators		
Category	Total	%
Arab ethnicity	26	9.8
Arab parent(s)	81	31.6
Lived in an Arabic-speaking country	42	15.9
Which languages are spoken in your home? (Arabic)	49	18.5
Do you speak Arabic at home? (Yes)	75	28.5
Is Arabic spoken in your home? (Yes; Used to be)	81	30.9

In terms of Arab parentage, 60 students indicated that both their parents spoke/speak Arabic, a further 21 saying that one of their parents speak/ spoke Arabic. 42 participants responded 'yes' to the question 'Have you ever lived in an Arabic speaking country (for longer than 3 months at a time)?'. The profiles of each of the participants who answered positively to one or more of the abovementioned items were analysed.

In order to qualify for HL status, a participant would need to have ticked one or more of the first three items. If more than two items in this primary category was ticked, this was regarded as an automatic qualification. If only one was ticked, this would need to be corroborated by a selection of one (or more) of the three secondary items. If only one (or more) of the secondary items were ticked, but no primary items, this was interpreted to mean that the participant had understood this to include the use of Arabic words either for homework or religious usage. 60% of the participants in this category had lived in an Arabic-speaking country, and 73.8% of these participants listed Arabic as one of the languages spoken in their homes.

As per the researcher's MSc study, attention is paid to the sub-divisions of these categories, since a NHL label does not necessarily mean the learner has no relationship with Arabic, in fact they may have some background factor which confers upon them an advantage over other NHLs when it comes to learning the language (Ramezanzadeh, 2015). A breakdown of participants' ethnic backgrounds is also provided in Table B.4.3 and illustrated in Figure B.4.1.

Table B.4.3. Participants' Ethnic Backgrounds				
	T1		T2	
Ethnicity	Frequency	Percentage	Frequency	Percentage
White English/ Scottish/ Irish	4	1.5	1	0.6
Any other white background	3	1.1	3	1.7
Asian Pakistani	62	23.4	46	26.3
Asian Bangladeshi	51	19.2	25	14.3
Asian Indian	31	18.8	20	11.4
Any other Asian background	12	4.5	9	5.1
Arab	26	9.8	20	11.4
Black African	44	16.6	30	17.1
Black Caribbean	0	0	1	0.6
Any other black background	2	0.8	3	1.7
Mixed White and Asian	7	2.6	3	1.7
Mixed White and Arab	0	0	1	0.6
Mixed White and Black African	2	0.8	1	0.6
Mixed White and Black Caribbean	0	0	0	0
Any other mixed background	21	7.8	12	6.9

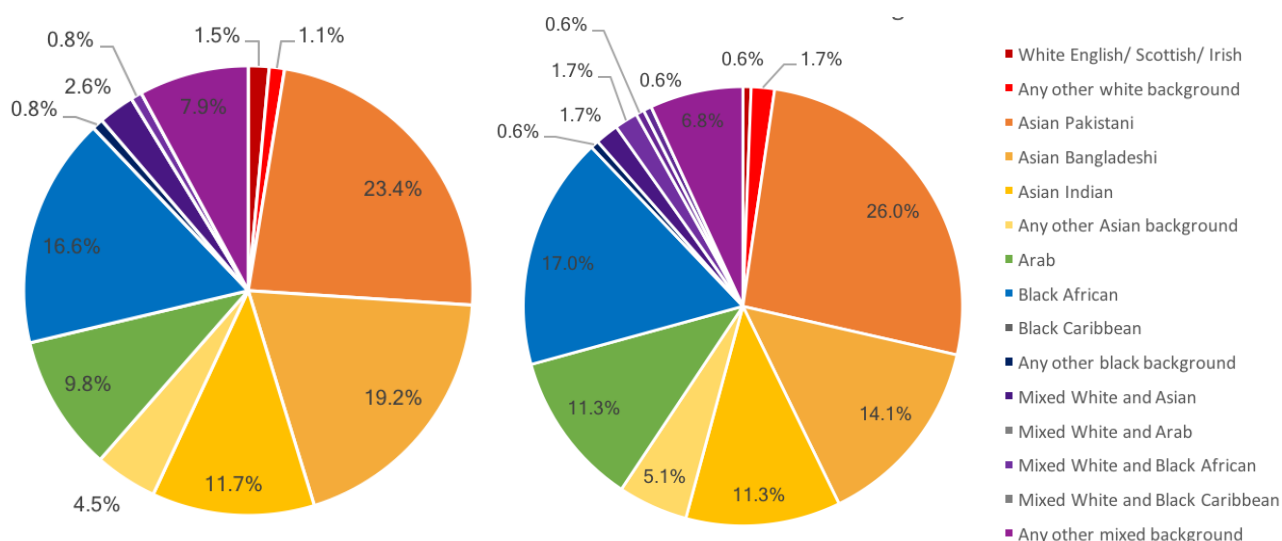


Figure B.4.1. Proportions of Participant Ethnic Backgrounds

In total, 39 students identified themselves of being of either Arab (27) or Mixed Arab (12) origin. Since the category of 'Arab' is a broad and disputed one, participants were asked to identify the country of origin from which their Arab heritage was derived. 10 different Arab countries were mentioned (Table B.4.4).

Table B.4.4. Country of Origin of Arab/ Mixed Arab participants		
Arab Country of Origin	Frequency	Percentage
Algeria	8	20.5
Iraq	3	7.7
Kuwait	2	5.1
Lebanon	2	5.1
Libya	5	12.8
Morocco	4	10.3
Qatar	1	2.6
Saudi Arabia	1	2.6
Sudan	1	2.6
Yemen	11	28.2
Unspecified	1	2.6

B.7. Consent Form & Information Sheets (Parents and Students)

DEPARTMENT OF EDUCATION

15 Norham Gardens
Oxford
OX2 6PY

Primary Researcher: Anna-Maria Ramezanzadeh – DPhil Candidate in Education
Email: anna-maria.ramezanzadeh@education.ox.ac.uk
Mob: 07714422709



Research study – Motivation and Engagement in the Arabic GCSE

Dear Parent/ Carer,

I am a research student at the University of Oxford, looking into the motivation and engagement of students studying the Arabic GCSE. Your child's school has kindly agreed to take part in this study, so I am writing to ask your permission for your child to be involved. This letter contains information about the study, please have a read before you decide. The students themselves will also be given information about the study and can choose whether or not to take part.

What does the study involve?

For this part of the study, students will be asked to complete a survey about their motivation towards studying the Arabic GCSE and a bit of information about their background, and two Arabic language activities. Students will be asked to fill these out twice, once this term, and once again later in the year. Completing these surveys and tasks will be done during one of their normal Arabic lessons, and shouldn't take more than 50 minutes.

Ethics

This research will be consistent with the strict ethical guidelines set by Oxford University. The study has been approved by the university's Research Ethics Committee as suitable for young people to take part in. Your son/daughter's school has also agreed to participate.

The data that we collect will be for research purposes only. No-one other than the researcher will have access to it. The data we collect is not connected with the school's assessment of your son/daughter's progress in any way. I will not share it with the school or its teachers. The data will be made anonymous and kept confidential, consistent with current UK law. All paper data will be stored safely in a locked filing cabinet in a locked university office, until such time as it has been converted to electronic format, after which it will be shredded. Electronic data will be stored in a dedicated computer folder, accessible only to the researcher using a password-protected log-in. Following standard university procedures, electronic data will be retained for as long as it may be of value for analysis, and usually for three years after publication of the study's findings. After this, it will be permanently deleted.

I plan to write up the findings in published articles, and present them at conferences. Any such reports or presentations will be anonymous: they will not identify students, teachers or the school. We will send a brief report of our findings to your child's school at the end of the project, and you are welcome to see this.

Does my son/daughter have to take part?

No. It is up to you and your son/daughter whether they take part. Your child will also be free to withdraw from the research at any point, without giving any reason. This will not affect their education in any way.

We hope that you and your son/daughter will be happy to take part in our study. However, if you prefer that s/he does **not** do so, please complete the reply slip below and return it to their Arabic teacher. Students themselves will also have the chance to opt out of the study using a similar slip. If you are happy for your child to take part, you do not need to do anything.

What if I have any questions or concerns?

If you have any questions or concerns about this project, please feel free to contact me. You can email me at the address on the front of this letter, write to me or telephone me using the contact details on the letter heading. You can also contact your son/daughter's Arabic teacher, who will pass your queries on to me.

If you are unhappy about any aspect of the study and wish to make a formal complaint, please contact the chair of the Research Ethics Committee at the University of Oxford (Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Hayes House, 75 George Street, Oxford, OX1 2BQ, UK). They will aim to resolve the matter as quickly as possible.

Thank you for reading this!
Best wishes,

Anna-Maria Ramezanzadeh

Reply slip

Research project: Motivation and Engagement in the Arabic GCSE

Tick the box if you **do not** wish your son/daughter to take part in the project, and complete the details. Please return the slip to your son/daughter's Arabic teacher.

You do not need to fill this in if you are happy for your son/daughter to take part in the project!

☐ **No** thank you - I would **not** like my son/daughter to take part in the research project.

Her/his name is _____

My name is _____

Signed _____

Date: _____

Later, the researcher will also sign this form:

Signed _____ Date: _____

DEPARTMENT OF EDUCATION

15 Norham Gardens
Oxford
OX2 6PY



Primary Researcher: Anna-Maria Ramezanzadeh – DPhil Candidate in Education
Email: anna-maria.ramezanzadeh@education.ox.ac.uk
Mob: 07714422709

Research Study – Motivation and Engagement in the Arabic GCSE

Dear Arabic Student,

My name is Anna-Maria and I am a student at the University of Oxford. I'm writing to you to invite you and other students in your Arabic class to take part in a study I am conducting. You can decide whether or not to take part. This letter contains some information about me and the study, to help you decide.

Who am I?

I am studying for my doctorate (PhD) at the Department of Education at the University of Oxford. I am interested in the experiences of students studying the Arabic GCSE. I am from a Muslim family (even though I have a strange name!) and was brought up in the UK. I studied Arabic from a young age and took the GCSE myself when I was in Year 10. I then went on to study Arabic at university for my degree. I've taught the language in an Islamic school to GCSE students, and am now teaching it to adults at the university.

What would you have to do?

If you agree to take part in this study, you will be asked to complete a survey about your motivation towards studying the Arabic GCSE and a bit of information about your background, and two activities. You will be asked to fill these out twice, once this term, and once again later in the year. You'll do this during one of your Arabic lessons so you won't have to spend any extra time on it outside of school. Your teacher will be asked to provide your current Arabic grade, and some basic background information which will help with the research.

My project has been approved by Oxford University as suitable for young people to take part in. I have done similar research in the past with GCSE students, who have usually found taking part in these projects interesting and enjoyable. I hope you will too!

What will I do with the information collected?

The information I collect will just be for research purposes. I will not tell your teacher how well you did in the Arabic task, or share your survey responses with them. I will not give you a mark or grade. All the information I collect will be confidential. When I transfer the data to a computer, I will give you a code number instead of using your name, so that all the information in my database is anonymous. At the end of the project, when I write or give presentations about what I have found out, I will not use your name or your school's name.

I will store electronic information safely on a secure computer, which only I can access, until I have finished analysing it. After this it will be permanently deleted. I will store the paper worksheets you complete in a locked filing cabinet in a locked university office until I have converted the information into electronic format. I will then shred the paper documents.

Do you have to take part?

No. It is up to you whether you would like to take part. I will not mind at all if you say no, and nor will your teacher. Even after the project has started, you can stop taking part at any time – just tell your teacher. This will not affect your progress in school in any way.

If you **don't** want to take part in this study, fill in the reply slip below and return it to your teacher, or just let your teacher know. If you are happy to take part, there is no need to do anything at this stage.

I will also write to your parents or carers to tell them about this project and to ask whether they are happy for you to take part. Please talk to them about it.

What if you have more questions?

If you have any questions about this project, please ask your teacher, or ask me if you see me in school. You can also contact me using the email address on the front of this sheet. I'd be happy to hear from you. The letter to your parents / carers will also explain other ways of getting in touch if you have any questions or concerns.

Thank you for reading this!
Best wishes,

Anna-Maria Ramezanzadeh

Reply slip

Research project: Motivation and Engagement in the Arabic GCSE

Tick the box if you **do not** wish to take part in the project. Then fill in your name and the date and return the slip to your teacher.

You do not need to fill this in if you are happy to take part in the project!

☐ **No** thank you - I would **not** like to take part in the research project.

Name: _____ Date: _____

B.8. Survey Instruction Letter & Protocol

DEPARTMENT OF EDUCATION

15 Norham Gardens
Oxford
OX2 6PY

Primary Researcher: Anna-Maria Ramezanzadeh – DPhil Candidate in Education
Email: anna-maria.ramezanzadeh@education.ox.ac.uk
Mob: 07714422709



Research Survey Protocols

Dear Teacher,

Thank you very much for offering to supervise the completion of the enclosed Surveys and Activity Booklets for my research study. You will find copies of these enclosed, one for each student and a spare in case.

Below are some instructions which will be helpful in ensuring this study runs smoothly and is in line with the other studies already conducted. To help with this, I have included a checklist on the next page. Please complete this as the study is carried out and return it with the completed surveys.

If you have any questions, please don't hesitate to contact me on the details above.

Thank you again,

Anna-Maria Ramezanzadeh

Research Study Supervision Protocol – Instructions

- 1) Ensure that only students who have consented are taking part. (I.e. only students who *did not* return consent forms).
- 2) Students should be seated at desks without any Arabic resources to hand.
- 3) Please read the following instructions to the class before giving out the Surveys:

This is a research study being run at Oxford University into the Arabic GCSE. The Researcher would like to know your honest thoughts, feelings, and opinions about learning the language. All the information you provide will be kept strictly anonymous and confidential. No one but the researcher and her supervisors will see these surveys. It is important that these surveys be completed individually. Please do not discuss your answers with your classmates or teacher whilst you are completing them. If you have any questions, please raise your hand and your teacher will answer any questions you might have. Once you have finished, please sit quietly and wait for the rest of the class to finish. It is not a competition! But the more information you can provide, the more useful it will be to the researcher. Any questions?

- 4) Hand out surveys and instruct students to fill out only the front cover.
- 5) Once all students have done this, they should open the survey, read the instructions, and begin. Please record the time the surveys were started on the checklist. Students should complete the surveys in silence.
- 6) Once all students have finished, please collect in the surveys and place them back in the envelop, ensuring that the contents of the booklets are not seen by anyone (including yourself).
- 7) Read the following instructions to the class before handing out the Activity Booklets:

This is a short Activity Booklet containing a couple of activities in Arabic. This is not a test. No one except the researcher will know your result. Your answers are simply to help the researcher with her analysis. This booklet must be completed individually. Do not share your answers or look at your classmates' answers. Please make sure to read the instructions for each question once you begin. If you find some of it difficult, don't worry, just do your best. If you have any questions, please raise your hand and ask your teacher.

- 8) Hand out Activity Booklets and instruct students to fill out only the front cover.
- 9) Once all students have done this, they should open the booklet, read the instructions, and begin. Please record the time the booklets were started on the checklist. Students should complete the booklets in silence.
- 10) Once all students have finished, please collect in the booklets and place them back in the envelop, ensuring that the contents of the booklets are not seen by anyone (including yourself).
- 11) Please record the time the final booklet was completed on the checklist.
- 12) Using the envelope provided, please return the completed Surveys and Activity Booklets by post to:

Anna-Maria Ramezanzadeh
C/o 8 St Omer Road, Cowley, Oxford OX4 3HB.

Research Study Supervision Protocol – Checklist

Participant Details	
Name of School:	
Date of Completion:	
Supervising Member of Staff:	
Number of Students:	Female: Male:
Year group of students:	

Procedure Checklist		
Procedure	✓/✗	Notes
Only students with consent taking part (i.e. those who did not return their forms)		
Explanation of study (Please see attached sheet)		
Students complete front cover only without opening the booklet		
All students start the survey at the same time		Time started:
All students completed survey (sitting in silence if they finish before others)		Time final student completed survey:
Explanation of Activity Book completion (Please see attached sheet)		
All students start at the same time		
All students completed Activity Booklet (sitting in silence if finished before others).		
Surveys completed independently and in silence.		
Activity Booklets completed independently and in silence.		
All booklets collected and packaged without contents being viewed by any individual (including class teacher) other than the student.		
Please give details of any questions students asked related to the surveys or Activity Booklets and answers provided by teacher/ supervisor (please continue on the back if necessary):		
Any other issues or comments raised during procedures? (Continue on the back if necessary)		

Signed by supervising member of staff: _____

B.9. Interview Analysis Coding Framework

Code (Sub-codes)	Description
Types of Arabic	References to use/ awareness of different types of Arabic
Reason for Arabic Choice	Reasons for taking Arabic
Arabic Future Self Images	-
AFS - Family	Communicating with family as part of AFS
AFS - General Communication	General ability to communicate in Arabic as part of AFS
AFS - Job	Use of Arabic in job as part of AFS
AFS - Just GCSE	No AFS other than getting the GCSE
AFS - No Arabic	No Arabic Future Self
AFS - Other	Any other AFS image
AFS - Religion	Using Arabic for religious reasons as part of AFS
AFS - Skill	Ability to use Arabic/ general knowledge of language
AFS - Travel	Use of Arabic in order to/ when travelling as part of AFS
Arabic Future Self - 'If'	Descriptions which include 'if' / other contingency/ hypothetical vocabulary
Arabic Future Self - 'When'	Descriptions which include 'when' / other expressions of certainty regarding FS image
Arabic Future Self - Changes	References to changes in Arabic future self-concepts
Current Self	Remarks about Arabic as part of current self-concept – comfort, confidence or current experience with Arabic learning
Current confidence	Remarks on current levels of confidence in Arabic
+ confidence	Increases in confidence in Arabic over the year
- confidence	Reductions in confidence in Arabic over the year
More effort needed	Participants referring to the need to make more effort themselves in order to do well
(Past) Experience with Arabic	Prior experiences with the Arabic language (e.g. living abroad, travelling abroad, Arabic community contact, Arabic family)

'Alimah course	References to mosque/ 'Alimah course study – details/ comparisons with GCSE
Perceptions of Arabic/ GCSE	-
'But'	Contradictory assessments of Arabic level/ learning experience
Difficulty	Descriptions of Arabic/ the GCSE as 'hard' or 'difficult'
Comparison to other languages	Comparisons between Arabic and other languages exposed to/ studied
New Spec	Comments on the new specification for the GCSE
GCSE Complaints	Complaints about the nature/ structure/ content of the GCSE
GCSE Stepping Stone	GCSE as a means to achieving another/ general future goal
Feelings	-
Apathy	Expressions of apathy towards Arabic/ the GCSE
Despondency	Feelings of despondency re. Arabic/ the GCSE - such as feelings of failure/ pointlessness
Dislike	Dislike of learning Arabic/ studying the GCSE
Enjoyment	References to enjoyment of language study
Negative	Negative emotions/ feeling towards Arabic/ GCSE
Positive	Positive emotions/ feeling towards Arabic/ GCSE
Future Self - General	-
Future Self - Clear	Expression of clear general future self-image (not related to Arabic)
Future Self - Unclear	No future self/ unclear general future self-image (not related to Arabic)
Ideal Self	-
IS - Desire	References to strength of desire to achieve Arabic future self-image
IS - Frequency	References to frequency of imaginations about Arabic future self-image
IS - Plausibility	References to likelihood of achieving Arabic future self-image
IS - Vividness	References to how clear/ vivid their Arabic future self-image is
Ought-to Self	-

Disappointment	References to feelings of disappointment from themselves/ others if learners do not succeed in Arabic
Force	References to being forced/ not having the choice to study Arabic
Compulsory	Arabic as a compulsory school option
Pressure	-
Exam Pressure	Pressure to do well in exams
Parents	Pressure from parents
Religion	Pressure from religion
School Pressure	Pressure from school/ teacher
Sixth Form College	Pressure to do well in order to get into Sixth Form/ college
Pressure Effects	Effects of pressure on learners
Self	Personal/ self-expectations/ pressure to do well
Consequences	Consequences for not doing well in Arabic (hypothetical)
Future plans affected	Future plans affected by not doing well in Arabic
Parents	Consequences from parents for not doing well
Personal Disappointment	Feelings of personal disappointment as a consequence for not doing well
'Waste'	References to not wanting to waste effort already invested in study/ threats of this
Parents	Any references to parents in relation to Arabic/ GCSE study
Priority	-
Arabic less	Arabic as less of a priority than other subjects
Arabic more	Arabic as more of a priority than other subjects
Religion Role	-
Religion - No role	Religion plays no role in learning Arabic
Religion Role - Afterlife	Learning Arabic in relation to afterlife (e.g. heaven)
Religion Role - Faith Effect	Effects of learning Arabic on faith/ religious practice/ relationship with God
Religion Role - Prophet	Arabic as language Prophet spoke as reason for learning it
Religion Role - Quran	Learning Arabic in order to read/ understand the Quran
'Reading' as sufficient	Necessity of only learning to read Arabic, rather than understand the Quran

Translations	References to translations of the Quran as impetus/ deterrent for learning
Task utility	Utility/ usefulness of classroom tasks in achieving Arabic Future Self images
Teacher Effect	Effects of teacher on learning of Arabic

C - FINDINGS

C.1. Missing Survey Data Statistics for Self-Concept Scales

Motivation Scales	T1			T2 (All Schools)			T2 (Adjusted Sample)		
Item	Missing	Percent	Valid N	Missing	Percent	Valid N	Missing	Percent	Valid N
Classical CS (Ability)	8	3.0%	261	96	35.7%	173	42	19.5%	173
Classical CS (Confidence)	6	2.2%	263	96	35.7%	173	41	19.1%	174
Classical CS (Self-Perception)	10	3.7%	259	96	35.7%	173	42	19.5%	173
Classical IS (Desire)	8	3.0%	261	95	35.3%	174	41	19.1%	174
Classical IS (Frequency)	8	3.0%	261	95	35.3%	174	41	19.1%	174
Classical IS (Plausibility)	14	5.2%	255	99	36.8%	170	45	20.9%	170
Classical IS (Vividness)	9	3.3%	260	96	35.7%	173	42	19.5%	173
Classical OTS (Consequences)	12	4.5%	257	96	35.7%	173	42	19.5%	173
Classical OTS (Disappointment)	11	4.1%	258	95	35.3%	174	41	19.1%	174
Classical OTS (Force)	8	3.0%	261	96	35.7%	173	42	19.5%	173
Classical OTS (Social Pressure)	8	3.0%	261	97	36.1%	172	42	19.5%	173
Dialect CS (Ability)	8	3.0%	261	99	36.8%	170	45	20.9%	170
Dialect CS (Confidence)	7	2.6%	262	98	36.4%	171	44	20.5%	171
Dialect CS (Self-Perception)	10	3.7%	259	100	37.2%	169	46	21.4%	169
Dialect IS (Desire)	6	2.2%	263	101	37.5%	168	46	21.4%	169
Dialect IS (Frequency)	7	2.6%	262	97	36.1%	172	43	20.0%	172
Dialect IS (Plausibility)	10	3.7%	259	98	36.4%	171	44	20.5%	171
Dialect IS (Vividness)	11	4.1%	258	98	36.4%	171	44	20.5%	171
Dialect OTS (Consequences)	9	3.3%	260	100	37.2%	169	45	20.9%	170
Dialect OTS (Disappointment)	8	3.0%	261	98	36.4%	171	44	20.5%	171
Dialect OTS (Force)	9	3.3%	260	99	36.8%	170	45	20.9%	170
Dialect OTS (Social Pressure)	7	2.6%	262	98	36.4%	171	44	20.5%	171
MSA CS (Ability)	12	4.5%	257	106	39.4%	163	51	23.7%	164
MSA CS (Confidence)	10	3.7%	259	102	37.9%	167	48	22.3%	167
MSA CS (Self-Perception)	11	4.1%	258	103	38.3%	166	49	22.8%	166
MSA IS (Desire)	10	3.7%	259	105	39.0%	164	51	23.7%	164
MSA IS (Frequency)	11	4.1%	258	105	39.0%	164	50	23.3%	165
MSA IS (Plausibility)	12	4.5%	257	103	38.3%	166	49	22.8%	166
MSA IS (Vividness)	9	3.3%	260	104	38.7%	165	50	23.3%	165
MSA OTS (Consequences)	11	4.1%	258	102	37.9%	167	48	22.3%	167
MSA OTS (Disappointment)	11	4.1%	258	103	38.3%	166	49	22.8%	166
MSA OTS (Force)	11	4.1%	258	103	38.3%	166	49	22.8%	166
MSA OTS (Social Pressure)	11	4.1%	258	103	38.3%	166	49	22.8%	166

C.2. Missing Survey Data Statistics for Self-Rated Proficiency Scales

Self-Rated Proficiency	T1			T2 (All Schools)			T2 (Adjusted Sample)		
Item	Missing	Percent	Valid N	Missing	Percent	Valid N	Missing	Percent	Valid N
Reading Words	5	1.9%	264	97	36.1%	172	43	20.0%	172
Word Reading (with Diacritics)	5	1.9%	264	100	37.2%	169	46	21.4%	169
Dialect (Listening Comprehension)	9	3.3%	260	101	37.5%	168	47	21.9%	168
Dialect (Speaking)	9	3.3%	260	101	37.5%	168	47	21.9%	168
MSA (Listening Comprehension)	5	1.9%	264	101	37.5%	168	47	21.9%	168
MSA (Reading Comprehension)	7	2.6%	262	99	36.8%	170	45	20.9%	170
MSA (Speaking)	6	2.2%	263	99	36.8%	170	45	20.9%	170
MSA (Writing)	6	2.2%	263	97	36.1%	172	43	20.0%	172
MSA v. Dialect	7	2.6%	262	100	37.2%	169	46	21.4%	169
Quran (Listening Comprehension)	7	2.6%	262	100	37.2%	169	46	21.4%	169
Quran (Reading Comprehension)	6	2.2%	263	100	37.2%	169	45	20.9%	170

C.3. Descriptive Statistics – Motivation Scales

	Variable (T1)	N	%	Miss.	%	Mean	Med	SD	Var.	Range	Min.	Max.	Skew.	Std Error	z- score	Kurt.	Std Error	z- score
MSA	CS Ability	250	92.6	20	7.4	2.81	3	1.445	2.089	5	1	6	0.421	0.154	2.734	-0.666	0.307	-2.169
	CS Confidence	250	92.6	20	7.4	3.16	3	1.527	2.333	5	1	6	0.123	0.154	0.799	-0.939	0.307	-3.059
	CS Self-Perception	250	92.6	20	7.4	3.08	3	1.573	2.476	5	1	6	0.303	0.154	1.968	-0.928	0.307	-3.023
	IS Desire	250	92.6	20	7.4	4.38	5	1.629	2.654	5	1	6	-0.652	0.154	-4.234	-0.723	0.307	-2.355
	IS Frequency	250	92.6	20	7.4	3.55	4	1.742	3.035	5	1	6	-0.099	0.154	-0.643	-1.230	0.307	-4.007
	IS Plausibility	250	92.6	20	7.4	3.78	4	1.541	2.375	5	1	6	-0.270	0.154	-1.753	-0.908	0.307	-2.958
	IS Vividness	250	92.6	20	7.4	3.71	4	1.633	2.665	5	1	6	-0.225	0.154	-1.461	-1.055	0.307	-3.436
	OTS Consequences	250	92.6	20	7.4	2.02	1	1.343	1.803	5	1	6	1.208	0.154	7.844	0.586	0.307	1.909
	OTS Disappointment	250	92.6	20	7.4	2.59	2	1.638	2.684	5	1	6	0.647	0.154	4.201	-0.778	0.307	-2.534
	OTS Force	250	92.6	20	7.4	1.87	1	1.294	1.673	5	1	6	1.471	0.154	9.552	1.461	0.307	4.759
	OTS Social Pressure	250	92.6	20	7.4	3.65	4	1.730	2.992	5	1	6	-0.219	0.154	-1.422	-1.204	0.307	-3.922
Dialect	CS Ability	250	92.6	20	7.4	3.39	3	1.390	1.933	5	1	6	0.057	0.154	0.370	-0.822	0.307	-2.678
	CS Confidence	250	92.6	20	7.4	3.35	3	1.551	2.406	5	1	6	0.113	0.154	0.734	-0.931	0.307	-3.033
	CS Self-Perception	250	92.6	20	7.4	3.33	3	1.533	2.351	5	1	6	0.163	0.154	1.058	-0.951	0.307	-3.098
	IS Desire	250	92.6	20	7.4	4.79	5	1.400	1.959	5	1	6	-0.884	0.154	-5.740	-0.138	0.307	-0.450
	IS Frequency	250	92.6	20	7.4	4.07	4	1.622	2.629	5	1	6	-0.448	0.154	-2.909	-0.949	0.307	-3.091
	IS Plausibility	250	92.6	20	7.4	4.08	4	1.435	2.058	5	1	6	-0.413	0.154	-2.682	-0.586	0.307	-1.909
	IS Vividness	250	92.6	20	7.4	4.12	4	1.523	2.320	5	1	6	-0.404	0.154	-2.623	-0.760	0.307	-2.476
	OTS Consequences	250	92.6	20	7.4	2.30	2	1.509	2.277	5	1	6	0.950	0.154	6.169	-0.121	0.307	-0.394
	OTS Disappointment	250	92.6	20	7.4	2.76	3	1.605	2.575	5	1	6	0.443	0.154	2.877	-0.942	0.307	-3.068
	OTS Force	250	92.6	20	7.4	2.17	2	1.450	2.103	5	1	6	1.146	0.154	7.442	0.391	0.307	1.274
	OTS Social Pressure	250	92.6	20	7.4	4.29	4.5	1.567	2.455	5	1	6	-0.638	0.154	-4.143	-0.617	0.307	-2.010
Classical	CS Ability	250	92.6	20	7.4	3.60	4	1.311	1.718	5	1	6	-0.186	0.154	-1.208	-0.507	0.307	-1.651
	CS Confidence	250	92.6	20	7.4	3.78	4	1.512	2.287	5	1	6	-0.260	0.154	-1.688	-0.851	0.307	-2.772
	CS Self-Perception	250	92.6	20	7.4	3.24	3	1.491	2.223	5	1	6	0.059	0.154	0.383	-0.934	0.307	-3.042
	IS Desire	250	92.6	20	7.4	4.78	5	1.292	1.668	5	1	6	-0.815	0.154	-5.292	-0.185	0.307	-0.603
	IS Frequency	250	92.6	20	7.4	3.89	4	1.559	2.429	5	1	6	-0.306	0.154	-1.987	-0.818	0.307	-2.664
	IS Plausibility	250	92.6	20	7.4	4.10	4	1.348	1.817	5	1	6	-0.471	0.154	-3.058	-0.433	0.307	-1.410
	IS Vividness	250	92.6	20	7.4	3.92	4	1.347	1.813	5	1	6	-0.229	0.154	-1.487	-0.725	0.307	-2.362
	OTS Consequences	250	92.6	20	7.4	2.82	3	1.473	2.170	5	1	6	0.383	0.154	2.487	-0.844	0.307	-2.749
	OTS Disappointment	250	92.6	20	7.4	3.03	3	1.666	2.774	5	1	6	0.360	0.154	2.338	-0.994	0.307	-3.238
	OTS Force	250	92.6	20	7.4	2.06	1	1.462	2.137	5	1	6	1.272	0.154	8.260	0.540	0.307	1.759
	OTS Social Pressure	250	92.6	20	7.4	4.37	5	1.573	2.474	5	1	6	-0.737	0.154	-4.786	-0.481	0.307	-1.567

	Variable (T2)	N	%	Miss.	%	Mean	Med	SD	Var.	Range	Min.	Max.	Skew.	Std Error	z-score	Kurt.	Std Error	z-score
MSA	CS Ability	167	61.9	103	38.1	2.89	3	1.519	2.306	5	1	6	0.425	0.188	2.261	-0.676	0.374	-1.807
	CS Confidence	167	61.9	103	38.1	3.19	3	1.579	2.493	5	1	6	0.181	0.188	0.963	-0.958	0.374	-2.561
	CS Self-Perception	167	61.9	103	38.1	2.91	3	1.631	2.661	5	1	6	0.357	0.188	1.899	-1.004	0.374	-2.684
	IS Desire	166	61.5	104	38.5	3.98	4	1.836	3.369	5	1	6	-0.440	0.188	-2.340	-1.178	0.375	-3.141
	IS Frequency	167	61.9	103	38.1	3.29	3	1.767	3.122	5	1	6	0.085	0.188	0.452	-1.321	0.374	-3.532
	IS Plausibility	167	61.9	103	38.1	3.50	4	1.707	2.914	5	1	6	-0.079	0.188	-0.420	-1.211	0.374	-3.238
	IS Vividness	167	61.9	103	38.1	3.51	4	1.856	3.444	5	1	6	-0.092	0.188	-0.489	-1.377	0.374	-3.682
	OTS Consequences	167	61.9	103	38.1	1.97	1	1.395	1.945	5	1	6	1.430	0.188	7.606	1.287	0.374	3.441
	OTS Disappointment	167	61.9	103	38.1	2.34	2	1.601	2.564	5	1	6	0.925	0.188	4.920	-0.372	0.374	-0.995
	OTS Force	166	61.5	104	38.5	2.09	1	1.545	2.386	5	1	6	1.365	0.188	7.261	0.716	0.375	1.909
Dialect	OTS Social Pressure	167	61.9	103	38.1	3.48	4	1.872	3.504	5	1	6	-0.056	0.188	-0.298	-1.438	0.374	-3.845
	CS Ability	171	63.3	99	36.7	3.37	3	1.455	2.118	5	1	6	-0.002	0.186	-0.011	-0.663	0.369	-1.797
	CS Confidence	171	63.3	99	36.7	3.36	3	1.536	2.360	5	1	6	-0.027	0.186	-0.145	-0.935	0.369	-2.534
	CS Self-Perception	172	63.7	98	36.3	3.19	3	1.623	2.635	5	1	6	0.191	0.185	1.032	-1.045	0.368	-2.840
	IS Desire	170	63.0	100	37.0	4.39	5	1.672	2.795	5	1	6	-0.802	0.186	-4.312	-0.487	0.370	-1.316
	IS Frequency	172	63.7	98	36.3	3.80	4	1.747	3.052	5	1	6	-0.270	0.185	-1.459	-1.190	0.368	-3.234
	IS Plausibility	171	63.3	99	36.7	3.94	4	1.593	2.538	5	1	6	-0.433	0.186	-2.328	-0.834	0.369	-2.260
	IS Vividness	171	63.3	99	36.7	3.95	4	1.642	2.697	5	1	6	-0.382	0.186	-2.054	-0.957	0.369	-2.593
	OTS Consequences	171	63.3	99	36.7	2.08	1	1.485	2.205	5	1	6	1.287	0.186	6.919	0.642	0.369	1.740
	OTS Disappointment	172	63.7	98	36.3	2.73	2	1.685	2.840	5	1	6	0.510	0.185	2.757	-1.054	0.368	-2.864
Classical	OTS Force	171	63.3	99	36.7	2.40	2	1.664	2.770	5	1	6	0.929	0.186	4.995	-0.419	0.369	-1.136
	OTS Social Pressure	171	63.3	99	36.7	4.18	5	1.651	2.726	5	1	6	-0.675	0.186	-3.629	-0.689	0.369	-1.867
	CS Ability	173	64.1	97	35.9	3.32	3	1.489	2.218	5	1	6	0.081	0.185	0.438	-0.953	0.367	-2.597
	CS Confidence	174	64.4	96	35.6	3.43	4	1.537	2.361	5	1	6	-0.065	0.184	-0.353	-0.979	0.366	-2.675
	CS Self-Perception	173	64.1	97	35.9	3.07	3	1.557	2.425	5	1	6	0.173	0.185	0.935	-0.987	0.367	-2.689
	IS Desire	174	64.4	96	35.6	4.40	5	1.648	2.716	5	1	6	-0.735	0.184	-3.995	-0.604	0.366	-1.650
	IS Frequency	174	64.4	96	35.6	3.74	4	1.729	2.991	5	1	6	-0.265	0.184	-1.440	-1.210	0.366	-3.306
	IS Plausibility	174	64.4	96	35.6	3.98	4	1.551	2.404	5	1	6	-0.498	0.184	-2.707	-0.678	0.366	-1.852
	IS Vividness	173	64.1	97	35.9	3.69	4	1.568	2.458	5	1	6	-0.222	0.185	-1.200	-0.963	0.367	-2.624
	OTS Consequences	173	64.1	97	35.9	2.40	2	1.425	2.032	5	1	6	0.767	0.185	4.146	-0.393	0.367	-1.071
	OTS Disappointment	174	64.4	96	35.6	2.76	2	1.739	3.025	5	1	6	0.501	0.184	2.723	-1.128	0.366	-3.082
	OTS Force	174	64.4	96	35.6	2.24	2	1.538	2.366	5	1	6	1.063	0.184	5.777	0.020	0.366	0.055
	OTS Social Pressure	173	64.1	97	35.9	4.23	4	1.604	2.574	5	1	6	-0.606	0.185	-3.276	-0.735	0.367	-2.003

C.4. Correlation Matrix – Self-Concept Scale Items

Motivation Scales T1

Spearman's rho (Non-Parametric Test of Correlation)

Motivation Scales T1	Classical CS (Ability)	Classical CS (Confidence)	Classical CS (Self-Perception)	Classical IS (Desire)	Classical IS (Frequency)	Classical IS (Plausibility)	Classical IS (Vividness)	Classical OTS (Consequences)	Classical OTS (Disappointment)	Classical OTS (Force)	Classical OTS (Social Pressure)	Dialect CS (Ability)	Dialect CS (Confidence)	Dialect CS (Self-Perception)	Dialect IS (Desire)	Dialect IS (Frequency)	Dialect IS (Plausibility)	Dialect IS (Vividness)	Dialect OTS (Consequences)	Dialect OTS (Disappointment)	Dialect OTS (Force)	Dialect OTS (Social Pressure)	MSA CS (Ability)	MSA CS (Confidence)	MSA CS (Current Self)	MSA IS (Desire)	MSA IS (Frequency)	MSA IS (Plausibility)	MSA IS (Vividness)	MSA OTS (Consequences)	MSA OTS (Disappointment)	MSA OTS (Force)	MSA OTS (Social Pressure)	
Spearman's rho (Non-Parametric Test of Correlation)																																		
Classical CS (Ability)	1.000																																	
Classical CS (Confidence)	.690**	1.000																																
Classical CS (Self-Perception)	.558**	.493**	1.000																															
Classical IS (Desire)	.143*	.174**	.135*	1.000																														
Classical IS (Frequency)	.379**	.376**	.374**	.623**	1.000																													
Classical IS (Plausibility)	.547**	.488**	.398**	.226**	.461**	1.000																												
Classical IS (Vividness)	.422**	.458**	.500**	.531**	.615**	.478**	1.000																											
Classical OTS (Consequences)	.163**	.165**	.125*	0.101	0.102	0.117	0.108	1.000																										
Classical OTS (Disappointment)	.220**	.214**	.174**	0.067	0.061	.139*	.157*	.570**	1.000																									
Classical OTS (Force)	0.048	-0.039	0.010	-.169**	-.170**	-0.081	-.123*	.303**	.462**	1.000																								
Classical OTS (Social Pressure)	.145*	.235**	.132*	.122*	0.093	.128*	.176**	.247**	.335**	.222**	1.000																							
Dialect CS (Ability)	.627**	.550**	.501**	.149*	.330**	.433**	.385**	.133*	.173**	0.069	0.060	1.000																						
Dialect CS (Confidence)	.606**	.671**	.465**	.140*	.366**	.464**	.422**	.197**	.206**	0.009	.177**	.714**	1.000																					
Dialect CS (Self-Perception)	.493**	.473**	.718**	0.110	.293**	.380**	.442**	.143*	.228**	0.009	.168**	.632**	.551**	1.000																				
Dialect IS (Desire)	.136*	.146*	0.081	.762**	.498**	.250**	.454**	0.061	0.044	-.182**	0.111	.180**	.220**	.180**	1.000																			
Dialect IS (Frequency)	.267**	.316**	.288**	.624**	.688**	.387**	.606**	0.083	0.051	-.139*	.162**	.316**	.425**	.281**	.621**	1.000																		
Dialect IS (Plausibility)	.463**	.492**	.354**	.325**	.458**	.685**	.506**	0.093	.173**	-0.064	.170**	.578**	.601**	.496**	.400**	.527**	1.000																	
Dialect IS (Vividness)	.400**	.437**	.331**	.488**	.539**	.455**	.676**	0.081	0.119	-0.097	.150*	.446**	.486**	.430**	.530**	.728**	.637**	1.000																
Dialect OTS (Consequences)	0.048	0.064	0.026	-0.052	-0.048	0.026	0.010	.577**	.554**	.410**	.266**	0.017	0.087	0.048	-0.024	0.024	0.011	-0.019	1.000															
Dialect OTS (Disappointment)	.138*	.161**	0.021	-0.026	-0.080	0.031	0.042	.512**	.675**	.513**	.332**	.214**	.224**	.126*	0.022	0.005	.162**	0.054	.625**	1.000														
Dialect OTS (Force)	0.020	-0.064	-0.098	-.184**	-.231**	-0.116	-.179**	.368**	.369**	.698**	.155*	0.088	0.021	-0.044	-.173**	-.125*	-0.098	-0.097	.462**	.564**	1.000													
Dialect OTS (Social Pressure)	0.098	.206**	0.079	0.090	0.065	0.110	.165**	.257**	.409**	.264**	.692**	.125*	.208**	.181**	.204**	.172**	.242**	.224**	.357**	.491**	.248**	1.000												
MSA CS (Ability)	.664**	.576**	.590**	0.105	.320**	.506**	.394**	0.103	.177**	-0.036	0.078	.614**	.593**	.587**	0.096	.240**	.482**	.359**	0.013	0.104	-0.014	0.077	1.000											
MSA CS (Confidence)	.582**	.612**	.484**	.146*	.386**	.488**	.443**	0.114	0.088	-0.024	0.107	.634**	.672**	.545**	.191**	.325**	.548**	.451**	-0.011	.127**	0.043	.185**	.742**	1.000										
MSA CS (Current Self)	.474**	.414**	.683**	.131*	.361**	.383**	.433**	0.086	.146*	-0.033	.187**	.534**	.502**	.716**	.177**	.325**	.455**	.377**	0.030	.152*	-0.021	.254**	.635**	.630**	1.000									
MSA IS (Desire)	.153*	.213**	.230**	.653**	.516**	.247**	.497**	0.067	0.025	-.151*	.147*	.217**	.248**	.222**	.696**	.568**	.336**	.509**	-0.028	0.026	-.157*	.219**	.216**	.314**	.332**	1.000								
MSA IS (Frequency)	.217**	.301**	.352**	.450**	.611**	.307**	.538**	.158*	0.074	-0.110	.166**	.280**	.354**	.312**	.436**	.662**	.411**	.551**	0.040	0.048	-0.088	.215**	.363**	.485**	.538**	.657**	1.000							
MSA IS (Plausibility)	.457**	.465**	.405**	.362**	.493**	.588**	.488**	0.098	.175**	-0.116	.184**	.491**	.499**	.481**	.380**	.482**	.733**	.571**	0.034	.131*	-0.109	.281**	.550**	.635**	.618**	.520**	.636**	1.000						
MSA IS (Vividness)	.444**	.477**	.432**	.431**	.613**	.522**	.668**	.128*	0.121	-.152*	.171**	.451**	.489**	.443**	.466**	.598**	.600**	.654**	0.049	0.090	-.166**	.273**	.481**	.611**	.627**	.615**	.741**	.731**	1.000					
MSA OTS (Consequences)	.140*	.136*	.147*	0.072	0.117	0.089	.158*	.496**	.523**	.425**	.256**	.189**	.149*	.137*	0.092	0.117	0.073	0.077	.638**	.596**	.485**	.306**	.208**	.195**	.226**	.147*	.242**	.152*	.198**	1.000				
MSA OTS (Disappointment)	0.058	0.118	0.045	-0.030	-0.052	0.037	0.052	.431**	.548**	.409**	.412**	.148*	0.121	0.120	0.034	-0.002	0.070	0.071	.538**	.687**	.411**	.510**	.146*	.158*	.173**	0.089	.167**	.165**	.135*	.644**	1.000			
MSA OTS (Force)	0.007	-0.044	0.027	-.132*	-.181**	-0.044	-0.050	.346**	.427**	.614**	.215**	0.050	0.035	0.093	-.131*	-0.112	-0.021	-0.029	.530**	.547**	.608**	.312**	0.070	0.060	.161**	-0.016	0.022	0.031	-0.007	.544**	.563**	1.000		
MSA OTS (Social Pressure)	.175**	.221**	.174**	.124*	0.114	0.099	.222**	.190**	.387**	.181**	.599**	.234**	.254**	.258**	.195**	.180**	.211**	.233**	.270**	.429**	.154*	.713**	.177**	.234**	.377**	.338**	.307**	.363**	.351**	.382**	.554**	.322**	1.000	

Significance indicators:

$p \leq .05 = *$ (light green)
 $p \leq .01 = **$ (darker green)

Motivation Scales T2

Spearman's rho (Non-Parametric Test of Correlation)

	Classical CS (Ability)	Classical CS (Confidence)	Classical CS (Self-Perception)	Classical IS (Desire)	Classical IS (Frequency)	Classical IS (Plausibility)	Classical IS (Vividness)	Classical OTS (Consequences)	Classical OTS (Disappointment)	Classical OTS (Force)	Classical OTS (Social Pressure)	Dialect CS (Ability)	Dialect CS (Confidence)	Dialect CS (Self-Perception)	Dialect IS (Desire)	Dialect IS (Frequency)	Dialect IS (Plausibility)	Dialect IS (Vividness)	Dialect OTS (Consequences)	Dialect OTS (Disappointment)	Dialect OTS (Force)	Dialect OTS (Social Pressure)	MSA CS (Ability)	MSA CS (Confidence)	MSA CS (Self-Perception)	MSA IS (Desire)	MSA IS (Frequency)	MSA IS (Plausibility)	MSA IS (Vividness)	MSA OTS (Consequences)	MSA OTS (Disappointment)	MSA OTS (Force)	MSA OTS (Social Pressure)
Classical CS (Ability)	1.000																																
Classical CS (Confidence)	.776**	1.000																															
Classical CS (Self-Perception)	.596**	.590**	1.000																														
Classical IS (Desire)	.334**	.412**	.386**	1.000																													
Classical IS (Frequency)	.447**	.493**	.513**	.737**	1.000																												
Classical IS (Plausibility)	.661**	.592**	.643**	.513**	.642**	1.000																											
Classical IS (Vividness)	.464**	.523**	.540**	.646**	.717**	.633**	1.000																										
Classical OTS (Consequences)	.165*	.212**	.254**	.180*	.153*	.193*	.225**	1.000																									
Classical OTS (Disappointment)	0.088	.164*	0.064	0.144	0.112	0.144	.197**	.489**	1.000																								
Classical OTS (Force)	-0.093	-0.049	-0.061	-0.090	-0.068	-0.147	-0.112	.317**	.377**	1.000																							
Classical OTS (Social Pressure)	.208**	.220**	0.127	.256**	.280**	.201**	.225**	.325**	.487**	.235**	1.000																						
Dialect CS (Ability)	.596**	.675**	.510**	.368**	.424**	.493**	.458**	.277**	.213**	-0.023	.199**	1.000																					
Dialect CS (Confidence)	.553**	.714**	.482**	.366**	.432**	.500**	.480**	.292**	.171*	-0.007	.179*	.762**	1.000																				
Dialect CS (Self-Perception)	.559**	.576**	.677**	.423**	.475**	.576**	.570**	.280**	.201**	-0.048	.193*	.712**	.665**	1.000																			
Dialect IS (Desire)	.297**	.336**	.332**	.749**	.662**	.457**	.555**	.222**	0.100	-0.086	.274**	.271**	.340**	.307**	1.000																		
Dialect IS (Frequency)	.327**	.432**	.379**	.630**	.660**	.485**	.640**	0.130	0.099	-0.065	.201**	.422**	.469**	.493**	.700**	1.000																	
Dialect IS (Plausibility)	.511**	.621**	.498**	.552**	.573**	.652**	.619**	.235**	.212**	-0.113	.276**	.620**	.710**	.709**	.544**	.553**	1.000																
Dialect IS (Vividness)	.388**	.468**	.383**	.587**	.636**	.527**	.670**	.270**	.180*	-0.081	.202**	.524**	.574**	.605**	.629**	.804**	.632**	1.000															
Dialect OTS (Consequences)	0.056	0.149	0.146	0.066	0.047	0.110	.180*	.479**	.531**	.431**	.263**	.269**	.228**	.257**	0.079	0.100	.183*	.205**	1.000														
Dialect OTS (Disappointment)	0.138	.198**	0.119	0.146	0.099	0.151	.261**	.479**	.735**	.371**	.469**	.287**	.258**	.246**	0.147	.169*	.270**	.265**	.705**	1.000													
Dialect OTS (Force)	-0.003	0.034	0.058	-0.023	-0.043	-0.009	-0.010	.402**	.391**	.661**	.191*	.156*	.168*	0.099	-0.025	0.030	-0.005	0.042	.566**	.519**	1.000												
Dialect OTS (Social Pressure)	.211**	.213**	0.092	.260**	.254**	.190*	.307**	.241**	.417**	0.130	.738**	.213**	.184*	.173*	.327**	.253**	.371**	.288**	.243**	.479**	.189*	1.000											
MSA CS (Ability)	.673**	.589**	.546**	.362**	.388**	.501**	.490**	.202**	0.074	-0.041	.167*	.669**	.582**	.592**	.269**	.301**	.484**	.375**	0.151	.183*	0.063	.185*	1.000										
MSA CS (Confidence)	.628**	.635**	.544**	.438**	.479**	.559**	.540**	.222**	0.050	-0.097	0.132	.676**	.674**	.587**	.333**	.396**	.571**	.495**	0.132	.163*	0.052	0.139	.807**	1.000									
MSA CS (Self-Perception)	.533**	.564**	.640**	.432**	.446**	.617**	.588**	.301**	0.150	-0.039	0.130	.621**	.587**	.735**	.331**	.488**	.554**	.552**	.207**	.210**	0.118	0.150	.662**	.668**	1.000								
MSA IS (Desire)	.314**	.380**	.332**	.668**	.573**	.434**	.575**	.285**	0.049	-0.069	0.153	.292**	.320**	.331**	.695**	.558**	.398**	.549**	0.112	.166*	0.051	.226**	.423**	.480**	.501**	1.000							
MSA IS (Frequency)	.336**	.451**	.443**	.539**	.619**	.492**	.661**	.219**	0.137	0.035	0.134	.463**	.471**	.463**	.564**	.667**	.443**	.607**	.235**	.204**	0.113	.159*	.453**	.510**	.621**	.726**	1.000						
MSA IS (Plausibility)	.581**	.656**	.492**	.547**	.562**	.652**	.586**	.222**	0.124	-0.162*	0.100	.682**	.634**	.563**	.490**	.514**	.643**	.578**	0.146	.160*	0.021	.218**	.674**	.745**	.700**	.610**	.673**	1.000					
MSA IS (Vividness)	.472**	.572**	.503**	.519**	.569**	.608**	.652**	.222**	0.088	-0.146	0.085	.574**	.533**	.544**	.428**	.540**	.545**	.627**	.168*	.175*	-0.016	.176*	.606**	.678**	.743**	.620**	.741**	.814**	1.000				
MSA OTS (Consequences)	0.095	.164*	.192*	.160*	0.059	.164*	.205**	.433**	.440**	.346**	0.128	.315**	.193*	.271**	0.064	.156*	.172*	.222**	.640**	.447**	.489**	.171*	.251**	.217**	.338**	.177*	.299**	.231**	.242**	1.000			
MSA OTS (Disappointment)	0.119	0.147	0.133	.192*	.156*	.233**	.235**	.392**	.658**	.337**	.346**	.273**	.218**	.211**	0.147	0.116	.192*	.233**	.619**	.698**	.465**	.316**	.241**	.214**	.331**	.237**	.309**	.275**	.243**	.686**	1.000		
MSA OTS (Force)	0.054	0.060	0.128	-0.002	-0.020	-0.022	0.026	.364**	.342**	.552**	.174*	.179*	0.091	0.111	-0.028	0.016	0.022	0.065	.502**	.480**	.645**	0.144	0.111	0.103	0.151	0.071	0.101	0.036	0.057	.590**	.559**	1.000	
MSA OTS (Social Pressure)	.282**	.264**	.198*	.253**	.281**	.285**	.341**	.298**	.400**	.183*	.547**	.395**	.256**	.309**	.182*	.227**	.206**	.284**	.326**	.459**	.294**	.604**	.382**	.303**	.469**	.360**	.390**	.407**	.439**	.406**	.545**	.357**	1.000

Significance indicators:

$p \leq .05 = *$ (light green)

$p \leq .01 = **$ (darker green)

T1		T2																																			
		Classical CS (Ability)	Classical CS (Confidence)	Classical CS (Self-Perception)	Classical IS (Desire)	Classical IS (Frequency)	Classical IS (Plausibility)	Classical IS (Vividness)	Classical OTS (Consequences)	Classical OTS (Disappointment)	Classical OTS (Force)	Classical OTS (Social Pressure)	Dialect CS (Ability)	Dialect CS (Confidence)	Dialect CS (Self-Perception)	Dialect IS (Desire)	Dialect IS (Frequency)	Dialect IS (Plausibility)	Dialect IS (Vividness)	Dialect OTS (Consequences)	Dialect OTS (Disappointment)	Dialect OTS (Force)	Dialect OTS (Social Pressure)	MSA CS (Ability)	MSA CS (Confidence)	MSA CS (Self-Perception)	MSA IS (Desire)	MSA IS (Frequency)	MSA IS (Plausibility)	MSA IS (Vividness)	MSA OTS (Consequences)	MSA OTS (Disappointment)	MSA OTS (Force)	MSA OTS (Social Pressure)			
Classical CS (Ability)		.647"	.565"	.422"	.192"	.249"	.420"	.287"	.151'	0.107	0.088	0.124	.454"	.429"	.348"	.172'	.155'	.357"	.223"	0.074	.135'	0.110	0.113	.559"	.490"	.334"	.164'	.186"	.389"	.299"	.167'	.140'	0.074	.186"			
Classical CS (Confidence)		.639"	.668"	.458"	.218"	.336"	.427"	.325"	0.122	0.023	-0.122	.167'	.493"	.552"	.437"	.220"	.261"	.458"	.315"	0.040	0.109	0.014	0.117	.544"	.552"	.399"	.238"	.262"	.473"	.422"	0.075	0.052	0.007	.170'			
Classical CS (Self-Perception)		.524"	.452"	.507"	.182"	.301"	.353"	.395"	0.090	0.103	0.010	.155'	.450"	.362"	.468"	0.129	.233"	.322"	.240"	.189"	.182"	.139'	0.094	.433"	.369"	.463"	0.117	.254"	.286"	.330"	.171'	.183'	0.124	.257"			
Classical IS (Desire)		.265"	.292"	.142'	.602"	.530"	.370"	.438"	.175'	.223"	0.020	.247"	.241"	.211"	.187"	.575"	.468"	.326"	.437"	0.068	.195"	0.042	.224"	.190"	.227"	.215"	.491"	.414"	.341"	.347"	.135'	.179"	0.033	.173'			
Classical IS (Frequency)		.360"	.420"	.253"	.430"	.530"	.385"	.463"	.129'	.155'	-0.022	.248"	.275"	.347"	.286"	.435"	.429"	.340"	.467"	0.086	.140'	0.034	.207"	.302"	.281"	.323"	.435"	.387"	.330"	.357"	0.065	.144'	0.018	.250"			
Classical IS (Plausibility)		.524"	.542"	.321"	.350"	.369"	.413"	.377"	0.105	0.081	-0.052	.203"	.410"	.448"	.322"	.319"	.312"	.502"	.313"	0.068	0.104	0.002	.256"	.391"	.367"	.259"	.253"	.229"	.413"	.327"	0.043	-0.003	-0.055	.138"			
Classical IS (Vividness)		.422"	.447"	.310"	.362"	.443"	.320"	.409"	.175'	0.112	-0.011	.210"	.378"	.382"	.340"	.371"	.402"	.409"	.414"	0.048	0.110	0.014	.234"	.272"	.307"	.300"	.339"	.363"	.310"	.348"	0.126	0.065	-0.022	.179"			
Classical OTS (Consequences)		.191"	.229"	0.109	.169"	.156'	0.121	.248"	.486"	.406"	.193"	.298"	.224"	.226"	.186"	0.024	0.119	.140'	.184"	.366"	.433"	.255"	.212"	.174'	.175'	.214"	.167"	.206"	.176"	.206"	.344"	.439"	.262"	.354"			
Classical OTS (Disappointment)		.196"	.183"	0.076	0.111	0.094	0.068	0.115	.389"	.555"	.258"	.389"	.303"	.269"	.244"	0.028	0.030	.230"	0.097	.383"	.515"	.325"	.356"	.243"	.197"	.181'	0.033	0.032	.165'	0.088	.306"	.449"	.275"	.331"			
Classical OTS (Force)		-0.091	0.016	-0.062	-0.084	-.144'	-.205"	-0.055	.217"	.319"	.448"	0.121	.134'	0.116	0.105	-.137'	-0.051	0.025	0.013	.436"	.361"	.393"	0.091	0.030	0.031	0.019	-.133'	-0.060	-0.078	-0.023	.290"	.286"	.410"	0.105			
Classical OTS (Social Pressure)		0.126	.152'	0.017	-0.012	0.045	0.020	0.057	.183"	.244"	0.045	.461"	0.126	.167"	0.067	-0.045	-0.012	0.058	0.036	.196"	.242"	.133'	.337"	0.099	0.058	0.076	-0.011	0.052	0.005	0.050	0.070	.157'	0.024	.346"			
Dialect CS (Ability)		.508"	.556"	.404"	.301"	.309"	.395"	.397"	.214"	.161'	0.088	.178"	.644"	.620"	.496"	.299"	.380"	.476"	.414"	.170'	.253"	.172'	.130'	.542"	.586"	.511"	.220"	.367"	.464"	.414"	.210"	.218"	.156'	.205"			
Dialect CS (Confidence)		.540"	.610"	.367"	.260"	.323"	.427"	.439"	.231"	.164'	-0.005	.210"	.587"	.637"	.488"	.256"	.354"	.484"	.445"	0.116	.227"	0.107	.170'	.527"	.596"	.489"	.275"	.378"	.487"	.456"	.194"	.162'	0.063	.242"			
Dialect CS (Self-Perception)		.505"	.482"	.558"	.254"	.313"	.435"	.454"	.160'	.198"	-0.110	.184"	.568"	.514"	.590"	.223"	.344"	.482"	.363"	.209"	.266"	0.100	.209"	.514"	.514"	.616"	.226"	.351"	.421"	.420"	.214"	.267"	.146'	.299"			
Dialect IS (Desire)		0.108	.164'	0.078	.531"	.430"	.309"	.413"	.169"	.251"	-0.051	.289"	.192"	.241"	.144'	.582"	.439"	.395"	.416"	0.111	.257"	0.063	.358"	.139"	.197"	.215"	.421"	.383"	.320"	.328"	0.105	.201"	-0.021	.217"			
Dialect IS (Frequency)		.302"	.371"	.181"	.445"	.488"	.442"	.499"	0.030	0.124	-0.095	.233"	.219"	.318"	.297"	.496"	.560"	.416"	.547"	0.117	.221"	0.030	.279"	.230"	.282"	.285"	.443"	.457"	.354"	.388"	0.037	0.128	-0.034	.180'			
Dialect IS (Plausibility)		.398"	.490"	.271"	.421"	.406"	.459"	.471"	0.103	.166'	-.140'	.153'	.475"	.557"	.395"	.371"	.459"	.574"	.492"	0.064	.167'	0.030	.205"	.388"	.446"	.397"	.309"	.356"	.511"	.447"	0.090	0.122	-0.085	.139'			
Dialect IS (Vividness)		.349"	.372"	.209"	.417"	.433"	.376"	.431"	0.075	.147'	-0.100	.151"	.390"	.455"	.398"	.404"	.474"	.469"	.529"	0.125	.207"	0.038	.233"	.334"	.414"	.287"	.360"	.406"	.414"	.389"	.156'	.166'	0.010	.192"			
Dialect OTS (Consequences)		-0.019	0.075	-0.021	-0.016	-0.074	0.000	-0.019	.421"	.332"	.218"	.205"	0.096	0.102	0.030	-0.064	-0.070	.148'	-0.010	.435"	.403"	.280"	.268"	0.046	0.028	0.045	0.004	-0.028	0.042	0.053	.351"	.333"	.263"	.228"			
Dialect OTS (Disappointment)		0.062	.152'	0.015	.153'	0.033	-0.024	0.105	.339"	.405"	.264"	.257"	.265"	.287"	.141'	0.042	0.086	.205"	.166'	.418"	.502"	.371"	.275"	.185"	.190"	.150'	0.081	0.110	.192"	.150'	.361"	.452"	.266"	.295"			
Dialect OTS (Force)		-.131'	-0.031	-0.102	-0.051	-.158'	-.199"	-0.070	.272"	.245"	.383"	0.009	0.110	.129"	0.048	-.130'	-0.073	-0.009	0.013	.348"	.307"	.393"	0.011	0.020	0.045	0.032	-0.060	0.042	-0.046	0.023	.319"	.331"	.388"	0.104			
Dialect OTS (Social Pressure)		0.053	.135'	-0.013	0.094	0.041	-0.045	0.092	.248"	.307"	0.032	.481"	.169"	.236"	0.063	0.074	0.056	.182"	0.094	.222"	.382"	.213"	.468"	.133'	.134'	0.069	0.039	0.007	0.100	0.078	0.106	.243"	0.090	.395"			
MSA CS (Ability)		.626"	.593"	.487"	.236"	.308"	.482"	.380"	.163'	0.057	-0.068	0.058	.599"	.557"	.497"	.140'	.277"	.409"	.265"	0.116	.152'	0.063	0.062	.657"	.577"	.483"	.143'	.231"	.459"	.359"	.172'	0.128	0.016	.202"			
MSA CS (Confidence)		.550"	.579"	.481"	.272"	.374"	.447"	.471"	.148'	0.018	-.155'	0.110	.601"	.563"	.500"	.270"	.423"	.458"	.442"	0.110	0.099	0.006	0.115	.559"	.612"	.500"	.280"	.390"	.533"	.472"	.132'	0.067	0.021	.219"			
MSA CS (Current Self)		.462"	.492"	.521"	.301"	.353"	.373"	.475"	.133'	0.118	-0.122	.208"	.501"	.450"	.511"	.282"	.412"	.382"	.399"	.170'	.196"	0.090	.198"	.485"	.433"	.533"	.284"	.446"	.435"	.485"	.170'	.215"	0.044	.349"			
MSA IS (Desire)		.191"	.296"	.203"	.563"	.440"	.251"	.400"	.263"	.175'	0.030	.245"	.252"	.245"	.234"	.554"	.471"	.286"	.450"	0.116	.224"	0.119	.276"	.251"	.297"	.333"	.577"	.436"	.417"	.420"	0.069	.190"	0.024	.294"			
MSA IS (Frequency)		.273"	.368"	.364"	.397"	.481"	.360"	.519"	.210"	0.112	-0.114	.199"	.264"	.298"	.366"	.356"	.487"	.277"	.476"	0.117	.160'	-0.007	.184"	.343"	.358"	.409"	.394"	.492"	.362"	.477"	0.050	0.121	-0.057	.305"			
MSA IS (Plausibility)		.430"	.551"	.434"	.445"	.473"	.467"	.469"	.242"	.181"	-0.086	.244"	.478"	.496"	.413"	.430"	.432"	.504"	.435"	0.126	.191"	0.058	.271"	.487"	.513"	.475"	.471"	.443"	.577"	.516"	.134'	.161'	0.011	.340"			
MSA IS (Vividness)		.414"	.539"	.349"	.397"	.505"	.412"	.481"	.134'	0.061	-.175'	.238"	.409"	.414"	.305"	.382"	.487"	.381"	.469"	0.063	0.123	-0.046	.207"	.414"	.472"	.381"	.429"	.464"	.501"	.529"	0.063	0.112	-0.048	.307"			
MSA OTS (Consequences)		0.079	.205"	0.085	.151'	0.124	0.071	0.096	.403"	.386"	.179"	.287"	.292"	.229"	.231"	.130'	.160'	.235"	.192"	.478"	.421"	.291"	.257"	.183'	0.125	.212"	.153'	.168'	.179"	.207"	.354"	.391"	.264"	.366"			
MSA OTS (Disappointment)		0.048	.151'	-0.031	0.101	0.062	0.018	0.052	.350"	.440"	.199"	.361"	.261"	.207"	0.111	0.085	0.038	.154'	0.124	.445"	.490"	.278"	.312"	.172'	0.111	.147'	0.084	0.059	0.116	.165'	.260"	.367"	.215"	.414"			
MSA OTS (Force)		-0.044	0.088	-0.004	-0.019	-0.123	-.161'	-0.003	.328"	.336"	.323"	0.092	.162'	.135'	0.122	-0.051	0.025	0.077	0.073	.376"	.360"	.337"	.130'	0.063	0.053	0.058	-0.038	0.026	0.006	0.062	.374"	.314"	.414"	.175'			
MSA OTS (Social Pressure)		.145'	.232"	0.078	.138'	0.093	0.049	0.118	.228"	.286"	0.039	.506"	.305"	.258"	.130'	.134'	0.082	.166'	.145'	.269"	.356"	.202"	.434"	.231"	.175'	.217"	.143'	.151'	.223"	.215"	.151'	.311"	.132'	.524"			

Significance indicators:

$p \leq .05 = *$ (light green)
 $p \leq .01 = **$ (darker green)

C.5. CFA Model Fit Indices for Language Varieties

Fit	Modification	N	Model Fit (Robust)	df	p	CFI	TLI	RMSEA	SRMR
			(as low as possible)						
MSA T1									
1b	with Plausibility as CS Item	248	176.902	41	0.000	0.888	0.85	0.129	0.096
1a	with Plausibility as IS Item		142.255	41	0.000	0.916	0.887	0.112	0.094
1c	without Plausibility	251	122.132	32	0.000	0.917	0.883	0.115	0.092
2	without Plausibility MSA.IS.Freq_T1 ~~ MSA.IS.Desire_T1		105.02	31	0.000	0.932	0.901	0.106	0.09
3	without Plausibility MSA.IS.Freq_T1 ~~ MSA.IS.Desire_T1 MSA.CS.Ab_T1 ~~ MSA.CS.Conf_T1		92.347	30	0.000	0.943	0.915	0.091	0.085
Dialect T1									
1a	with Plausibility as IS Item	247	128.729	41	0.000	0.925	0.9	0.097	0.079
1b	with Plausibility as CS Item		124.54	41	0.000	0.928	0.904	0.095	0.076
1c	without Plausibility	250	89.649	32	0.000	0.944	0.921	0.086	0.068
2	without Plausibility D.IS.Freq_T1 ~~ D.IS.Desire_T1		75.718	31	0.000	0.957	0.937	0.077	0.064
Classical T1									
1a	with Plausibility as IS Item	235	128.389	41	0.000	0.895	0.86	0.1	0.088
1b	with Plausibility as CS Item		101.633	41	0.000	0.927	0.902	0.083	0.083
1c	without Plausibility	243	86.45	32	0.000	0.93	0.901	0.086	0.082
2	without Plausibility Cl.IS.Freq_T1~~Cl.IS.Des_T1		65.015	31	0.000	0.958	0.936	0.069	0.074
MSA T2									
1	without Plausibility	158	114.049	32	0.000	0.896	0.853	0.142	0.088
2	without Plausibility MSA.CS.Ab_T2~~MSA.CS.Conf_T2		93.29	31	0.000	0.923	0.888	0.125	0.079
3	without Plausibility MSA.CS.Ab_T2~~MSA.CS.Conf_T2 MSA.IS.Desire_T2~~MSA.IS.Freq_T2		70.802	30	0.000	0.949	0.924	0.102	0.078
4	without Plausibility MSA.CS.Ab_T2 ~~ MSA.CS.Conf_T2 MSA.IS.Desire_T2~~MSA.IS.Freq_T2 MSA.OTS.Cons_T2~~MSA.OTS.SocPress_T2		64.295	29	0.000	0.95	0.922	0.095	0.069
Dialect T2									
1	without Plausibility	160	52.728	32	0.012	0.969	0.956	0.073	0.064
2	without Plausibility D.IS.Freq_T2 ~~ D.IS.Desire_T2		42.171	31	0.087	0.983	0.976	0.054	0.059
Classical T2									
1	without Plausibility	166	63.753	32	0.000	0.95	0.93	0.083	0.08
2	without Plausibility Cl.CS.Ab T2 ~~ Cl.CS.Conf T2		57.255	31	0.003	0.959	0.941	0.076	0.075

C.6. CFA Model Fit Indices for Three-Factor Structure

Fit	Modification	N	Model Fit (Robust) (as low as possible)	df	p	CFI (good if >0.9, great if >0.95)	TLI	RMSEA (good if <0.1, great if <0.05)	SRMR
CS T1									
1	Intervariant items correlated	250	26.458	15	0.033	0.992	0.982	0.056	0.027
IS T1									
1	Intervariant items correlated	250	24.788	15	0.053	0.993	0.983	0.056	0.025
OTS T1									
1	Intervariant items correlated	250	69.394	39	0.002	0.983	0.971	0.055	0.040
CS T2									
1	Intervariant items correlated	152	15.641	15	0.406	0.999	0.998	0.019	0.048
IS T2									
1	Intervariant items correlated	152	12.606	15	0.633	1.000	1.007	0.000	0.024
OTS T2									
1	Intervariant items correlated	152	44.81	39	0.241	0.994	0.990	0.034	0.052

C.7. CFA Three-Factor Structure Factor Loadings and Variance and Covariance Coefficients

	T1						T2					
	MSA (Fit 4)		Dialect (Fit 2)		Classical (Fit 2)		MSA (Fit 4)		Dialect (Fit 2)		Classical (Fit 2)	
Variable	Factor Loading	Error	Factor Loading	Error	Factor Loading	Error	Factor Loading	Error	Factor Loading	Error	Factor Loading	Error
<i>CS Ability</i>	0.9	0.18	0.89	0.22	0.87	0.24	0.88	0.22	0.92	0.15	0.88	0.23
<i>CS Confidence</i>	0.84	0.29	0.78	0.39	0.76	0.42	0.89	0.21	0.85	0.28	0.91	0.18
<i>CS Self-concept</i>	0.69	0.52	0.71	0.5	0.67	0.55	0.74	0.46	0.76	0.42	0.67	0.54
Current Self	0.93	0.14	0.94	0.12	0.88	0.22	0.91	0.16	0.92	0.16	0.86	0.25
<i>IS Desire</i>	0.76	0.42	0.73	0.5	0.71	0.46	0.79	0.37	0.78	0.39	0.8	0.35
<i>IS Frequency</i>	0.88	0.23	0.85	0.18	0.9	0.28	0.93	0.14	0.91	0.17	0.89	0.2
<i>IS Vividness</i>	0.85	0.28	0.74	0.41	0.77	0.45	0.78	0.39	0.83	0.31	0.82	0.33
Ideal Self	0.86	0.25	0.93	0.13	0.95	0.1	0.83	0.31	0.94	0.12	0.94	0.11
<i>OTS Disappointment</i>	0.64	0.59	0.59	0.65	0.56	0.69	0.7	0.51	0.63	0.6	0.54	0.72
<i>OTS Force</i>	0.88	0.23	0.92	0.16	0.89	0.21	0.89	0.21	0.86	0.27	0.73	0.46
<i>OTS Social Pressure</i>	0.73	0.47	0.71	0.5	0.59	0.66	0.76	0.42	0.85	0.29	0.65	0.57
<i>OTS Disappointment</i>	0.57	0.67	0.57	0.68	0.46	0.79	0.64	0.6	0.49	0.76	0.57	0.68
Ought-To Self	0.89	0.2	0.98	0.05	0.92	0.15	0.81	0.34	0.99	0.02	0.87	0.25

CURRENT SELF			T1				T2			
Latent Variable		Obs Variable	Estimate	St. Error	z-value	p-value	Estimate	St. Error	z-value	p-value
CS_MSA	==	MSA.CS.Ab	1.000	0.000	NA	NA	1.000	0.000	NA	NA
CS_MSA	==	MSA.CS.Conf	0.986	0.059	16.668	0.000	1.058	0.064	16.663	0.000
CS_MSA	==	MSA.CS.SelfPerc	0.815	0.066	12.385	0.000	0.865	0.083	10.382	0.000
CS_Cl	==	Cl.CS.Ab	1.000	0.000	NA	NA	1.000	0.000	NA	NA
CS_Cl	==	Cl.CS.Conf	1.012	0.076	13.294	0.000	1.082	0.075	14.525	0.000
CS_Cl	==	Cl.CS.SelfPerc	0.870	0.083	10.465	0.000	0.798	0.096	8.346	0.000
CS_D	==	D.CS.Ab	1.000	0.000	NA	NA	1.000	0.000	NA	NA
CS_D	==	D.CS.Conf	0.985	0.071	13.933	0.000	0.989	0.062	15.910	0.000
CS_D	==	D.CS.SelfPerc	0.869	0.074	11.780	0.000	0.881	0.098	8.976	0.000
Current_Self	==	CS_MSA	1.000	0.000	NA	NA	1.000	0.000	NA	NA
Current_Self	==	CS_Cl	0.833	0.060	13.953	0.000	0.908	0.067	13.650	0.000
Current_Self	==	CS_D	0.953	0.071	13.457	0.000	0.981	0.088	11.189	0.000
Covariances										
MSA.CS.Ab	~~	Cl.CS.Ab	-0.010	0.064	-0.154	0.877	0.169	0.086	1.964	0.050
MSA.CS.Ab	~~	D.CS.Ab	-0.108	0.067	-1.615	0.106	0.043	0.066	0.654	0.513
Cl.CS.Ab	~~	D.CS.Ab	-0.011	0.060	-0.179	0.858	-0.073	0.067	-1.084	0.279
MSA.CS.Conf	~~	Cl.CS.Conf	0.214	0.087	2.465	0.014	0.037	0.103	0.356	0.721
MSA.CS.Conf	~~	D.CS.Conf	0.244	0.107	2.275	0.023	0.200	0.116	1.729	0.084
Cl.CS.Conf	~~	D.CS.Conf	0.358	0.103	3.482	0.000	0.247	0.112	2.196	0.028
MSA.CS.SelfPerc	~~	Cl.CS.SelfPerc	0.631	0.150	4.204	0.000	0.485	0.192	2.528	0.011
MSA.CS. SelfPerc	~~	D.CS.SelfPerc	0.648	0.149	4.351	0.000	0.521	0.198	2.634	0.008
Cl.CS. SelfPerc	~~	D.CS.SelfPerc	0.670	0.144	4.645	0.000	0.569	0.201	2.835	0.005
Variances										
MSA.CS.Ab			0.376	0.102	3.676	0.000	0.513	0.106	4.836	0.000
MSA.CS.Conf			0.681	0.122	5.576	0.000	0.517	0.151	3.419	0.001
MSA.CS. SelfPerc			1.243	0.175	7.108	0.000	1.123	0.227	4.938	0.000
Cl.CS.Ab			0.410	0.089	4.588	0.000	0.496	0.121	4.100	0.000
Cl.CS.Conf			0.959	0.132	7.251	0.000	0.409	0.122	3.368	0.001
Cl.CS. SelfPerc			1.220	0.162	7.534	0.000	1.248	0.213	5.867	0.000
D.CS.Ab			0.414	0.097	4.258	0.000	0.310	0.104	2.991	0.003
D.CS.Conf			0.948	0.152	6.251	0.000	0.642	0.150	4.292	0.000
D.CS. SelfPerc			1.128	0.181	6.243	0.000	0.978	0.235	4.155	0.000
CS_MSA			0.232	0.077	3.014	0.003	0.290	0.092	3.137	0.002
CS_Cl			0.285	0.089	3.205	0.001	0.413	0.130	3.188	0.001
CS_D			0.179	0.083	2.151	0.031	0.276	0.134	2.056	0.040
Current_Self			1.467	0.196	7.498	0.000	1.485	0.238	6.254	0.000

IDEAL SELF			T1				T2			
Latent Variable		Obs Variable	Estimate	St. Error	z-value	p-value	Estimate	St. Error	z-value	p-value
IS_MSA	=~	MSA.IS.Desire	1.000	0.000	NA	NA	1.000	0.000	NA	NA
IS_MSA	=~	MSA.IS.Freq	1.230	0.098	12.616	0.000	1.139	0.099	11.545	0.000
IS_MSA	=~	MSA.IS.Vivid	1.114	0.089	12.580	0.000	0.991	0.096	10.266	0.000
IS_Cl	=~	Cl.IS.Des	1.000	0.000	NA	NA	1.000	0.000	NA	NA
IS_Cl	=~	Cl.IS.Freq	1.408	0.111	12.685	0.000	1.188	0.079	14.946	0.000
IS_Cl	=~	Cl.IS.Vivid	1.074	0.104	10.374	0.000	0.979	0.085	11.458	0.000
IS_D	=~	D.IS.Desire	1.000	0.000	NA	NA	1.000	0.000	NA	NA
IS_D	=~	D.IS.Freq	1.509	0.144	10.482	0.000	1.187	0.107	11.136	0.000
IS_D	=~	D.IS.Vivid	1.197	0.106	11.294	0.000	1.019	0.087	11.656	0.000
Ideal_Self	=~	IS_MSA	1.000	0.000	NA	NA	1.000	0.000	NA	NA
Ideal_Self	=~	IS_Cl	0.828	0.064	13.007	0.000	1.026	0.105	9.752	0.000
Ideal_Self	=~	IS_D	0.848	0.068	12.386	0.000	1.027	0.111	9.247	0.000
Covariances										
MSA.IS.Desire	~~	Cl.IS.Des	0.462	0.090	5.155	0.000	0.438	0.143	3.059	0.002
MSA.IS.Desire	~~	D.IS.Desire	0.525	0.104	5.053	0.000	0.614	0.156	3.930	0.000
Cl.IS.Des	~~	D.IS.Desire	0.513	0.081	6.346	0.000	0.451	0.144	3.125	0.002
MSA.IS.Freq	~~	Cl.IS.Freq	-0.018	0.136	-0.135	0.893	-0.078	0.114	-0.691	0.490
MSA.IS.Freq	~~	D.IS.Freq	0.134	0.138	0.973	0.331	0.096	0.103	0.931	0.352
Cl.IS.Freq	~~	D.IS.Freq	0.053	0.125	0.420	0.675	-0.082	0.131	-0.622	0.534
MSA.IS.Vivid	~~	Cl.IS.Vivid	0.307	0.086	3.572	0.000	0.258	0.140	1.845	0.065
MSA.IS.Vivid	~~	D.IS.Vivid	0.303	0.099	3.053	0.002	0.322	0.167	1.930	0.054
Cl.IS.Vivid	~~	D.IS.Vivid	0.381	0.111	3.427	0.001	0.206	0.112	1.834	0.067
Variances										
MSA.IS.Desire			1.134	0.139	8.136	0.000	1.199	0.226	5.300	0.000
MSA.IS.Freq			0.709	0.190	3.738	0.000	0.426	0.170	2.502	0.012
MSA.IS.Vivid			0.728	0.130	5.594	0.000	1.306	0.276	4.731	0.000
Cl.IS.Des			0.758	0.092	8.276	0.000	0.911	0.147	6.184	0.000
Cl.IS.Freq			0.681	0.147	4.628	0.000	0.586	0.116	5.041	0.000
Cl.IS.Vivid			0.812	0.112	7.241	0.000	0.796	0.172	4.636	0.000
D.IS.Desire			0.954	0.109	8.724	0.000	1.057	0.188	5.608	0.000
D.IS.Freq			0.480	0.186	2.584	0.010	0.477	0.181	2.630	0.009
D.IS.Vivid			0.932	0.168	5.559	0.000	0.776	0.148	5.233	0.000
IS_MSA			0.388	0.118	3.280	0.001	0.628	0.184	3.407	0.001
IS_Cl			0.087	0.047	1.864	0.062	0.179	0.121	1.478	0.139
IS_D			0.121	0.038	3.148	0.002	0.196	0.110	1.783	0.075
Ideal_Self			1.149	0.195	5.908	0.000	1.412	0.283	4.984	0.000

OUGHT-TO SELF			T1				T2			
Latent Variable		Obs Variable	Estimate	St. Error	z-value	p-value	Estimate	St. Error	z-value	p-value
OTS_MSA	==	MSA.OTS.For	1.000	0.000	NA	NA	1.000	0.000	NA	NA
OTS_MSA	==	MSA.OTS.Diss	1.737	0.211	8.229	0.000	1.349	0.214	6.312	0.000
OTS_MSA	==	MSA.OTS.Cons	1.190	0.155	7.672	0.000	1.016	0.150	6.783	0.000
OTS_MSA	==	MSA.OTS.SocPress	1.192	0.172	6.929	0.000	1.133	0.204	5.556	0.000
OTS_Cl	==	Cl.OTS.Forc	1.000	0.000	NA	NA	1.000	0.000	NA	NA
OTS_Cl	==	Cl.OTS.Diss	1.817	0.256	7.089	0.000	1.547	0.323	4.791	0.000
OTS_Cl	==	Cl.OTS.Cons	1.053	0.150	7.038	0.000	1.117	0.209	5.356	0.000
OTS_Cl	==	Cl.OTS.SocPress	0.888	0.141	6.298	0.000	1.076	0.226	4.773	0.000
OTS_D	==	D.OTS.Force	1.000	0.000	NA	NA	1.000	0.000	NA	NA
OTS_D	==	D.OTS.Diss	1.728	0.227	7.602	0.000	1.419	0.200	7.089	0.000
OTS_D	==	D.OTS.Cons	1.258	0.130	9.656	0.000	1.260	0.147	8.561	0.000
OTS_D	==	D.OTS.SocPress	1.056	0.163	6.491	0.000	0.780	0.157	4.967	0.000
OughtTo_SelfT1	==	OTS_MSA	1.000	0.000	NA	NA	1.000	0.000	NA	NA
OughtTo_SelfT1	==	OTS_Cl	1.021	0.131	7.789	0.000	0.852	0.148	5.744	0.000
OughtTo_SelfT1	==	OTS_D	1.124	0.148	7.586	0.000	1.199	0.137	8.728	0.000
Covariances										
MSA.OTS.For	~~	Cl.OTS.Forc	0.611	0.154	3.981	0.000	0.753	0.189	3.991	0.000
MSA.OTS.For	~~	D.OTS.Force	0.535	0.146	3.658	0.000	0.776	0.224	3.470	0.001
Cl.OTS.Forc	~~	D.OTS.Force	0.901	0.151	5.952	0.000	0.984	0.230	4.276	0.000
MSA.OTS.Diss	~~	Cl.OTS.Diss	-0.188	0.112	-1.679	0.093	0.417	0.164	2.545	0.011
MSA.OTS.Diss	~~	D.OTS.Diss	0.017	0.121	0.139	0.890	0.216	0.133	1.626	0.104
Cl.OTS.Diss	~~	D.OTS.Diss	-0.086	0.123	-0.706	0.480	0.592	0.215	2.759	0.006
MSA.OTS.Cons	~~	Cl.OTS.Cons	0.294	0.105	2.790	0.005	0.238	0.150	1.586	0.113
MSA.OTS.Cons	~~	D.OTS.Cons	0.358	0.099	3.623	0.000	0.408	0.166	2.452	0.014
Cl.OTS.Cons	~~	D.OTS.Cons	0.422	0.106	3.987	0.000	0.128	0.166	0.771	0.441
MSA.OTS.SocPress	~~	Cl.OTS.SocPress	1.009	0.145	6.948	0.000	0.803	0.197	4.081	0.000
MSA.OTS.SocPress	~~	D.OTS.SocPress	1.141	0.142	8.033	0.000	1.037	0.198	5.234	0.000
Cl.OTS.SocPress	~~	D.OTS.SocPress	1.141	0.142	8.014	0.000	1.199	0.188	6.386	0.000
Variances										
MSA.OTS.For			0.996	0.169	5.896	0.000	1.145	0.232	4.936	0.000
MSA.OTS.Diss			0.624	0.199	3.140	0.002	0.517	0.158	3.275	0.001
MSA.OTS.Cons			0.843	0.132	6.398	0.000	0.806	0.187	4.315	0.000
MSA.OTS.SocPress			1.963	0.185	10.610	0.000	2.041	0.240	8.498	0.000
Cl.OTS.Forc			1.476	0.199	7.405	0.000	1.692	0.249	6.806	0.000
Cl.OTS.Diss			0.569	0.195	2.917	0.004	1.431	0.319	4.487	0.000
Cl.OTS.Cons			1.409	0.136	10.366	0.000	1.168	0.223	5.247	0.000
Cl.OTS.SocPress			1.939	0.148	13.061	0.000	1.681	0.192	8.739	0.000
D.OTS.Force			1.337	0.158	8.469	0.000	1.587	0.295	5.389	0.000
D.OTS.Diss			0.416	0.139	3.004	0.003	0.776	0.219	3.540	0.000
D.OTS.Cons			1.121	0.179	6.266	0.000	0.666	0.193	3.449	0.001
D.OTS.SocPress			1.666	0.142	11.750	0.000	2.023	0.203	9.942	0.000
OTS_MSA			0.138	0.053	2.592	0.010	0.366	0.138	2.660	0.008
OTS_Cl			0.100	0.038	2.615	0.009	0.175	0.095	1.835	0.067
OTS_D			0.035	0.033	1.048	0.294	0.019	0.077	0.245	0.807
OughtTo_SelfT1			0.541	0.136	3.985	0.000	0.715	0.267	2.681	0.007

C.8. Reliability Scores – Self-Concept Scale Items

Construct	Items	N	N of Items	Cronbach's Alpha	Cronbach's Alpha (Standardized Items)	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
T1										
Ideal Self										
Classical Ideal Self	Cl.IS.Plaus	260	4	0.797	0.796	12.63	12.905	0.473	0.29	0.81
	Cl.IS.Freq					12.84	9.835	0.724	0.541	0.685
	Cl.IS.Des					11.94	12.351	0.578	0.439	0.762
	Cl.IS.Vivid					12.78	11.307	0.679	0.461	0.713
Dialect Ideal Self	D.IS.Vivid	261	4	0.847	0.846	12.96	13.706	0.77	0.609	0.767
	D.IS.Freq					13	13.338	0.74	0.572	0.781
	D.IS.Plaus					12.99	15.546	0.621	0.434	0.831
	D.IS.Desire					12.29	15.89	0.613	0.405	0.834
MSA Ideal Self	MSA.IS.Vivid	259	4	0.886	0.886	11.7	18.093	0.817	0.68	0.827
	MSA.IS.Freq					11.89	17.74	0.779	0.616	0.843
	MSA.IS.Plaus					11.68	19.823	0.727	0.565	0.862
	MSA.IS.Desire					11.02	19.643	0.685	0.48	0.877
Current Self										
Classical Current Self	Cl.CS.Ab	259	3	0.802	0.806	7.04	6.882	0.728	0.548	0.655
	Cl.CS.Current					7.36	7.062	0.563	0.327	0.819
	Cl.CS.Conf					6.86	6.371	0.665	0.502	0.711
Dialect Current Self	D.CS.Ab	260	3	0.831	0.834	6.65	7.463	0.762	0.588	0.702
	D.CS.Current					6.7	7.531	0.629	0.413	0.827
	D.CS.Conf					6.69	7.048	0.688	0.519	0.769
MSA Current Self	MSA.CS.Current	256	3	0.862	0.864	5.97	7.889	0.679	0.462	0.864
	MSA.CS.Conf					5.89	7.619	0.767	0.616	0.78
	MSA.CS.Ab					6.23	8.008	0.775	0.622	0.775
Ought-to Self										
Classical OT Self	Cl.OTS.Cons	259	4	0.71	0.708	9.48	13.142	0.509	0.316	0.64
	Cl.OTS.Diss					9.27	10.882	0.64	0.432	0.549
	Cl.OTS.Forc					10.22	13.558	0.463	0.238	0.667
	Cl.OTS.SocPress					7.93	13.876	0.386	0.155	0.713
Dialect OT Self	D.OTS.Force	260	4	0.782	0.782	9.31	14.453	0.569	0.371	0.739
	D.OTS.Diss					8.7	12.012	0.728	0.533	0.651
	D.OTS.SocPress					7.2	14.74	0.456	0.246	0.797
	D.OTS.Cons					9.19	13.7	0.615	0.415	0.716
MSA OT Self	MSA.OTS.SocPress	257	4	0.791	0.798	6.49	13.196	0.51	0.301	0.797
	MSA.OTS.Cons					8.11	14.319	0.641	0.452	0.725
	MSA.OTS.Diss					7.55	11.686	0.743	0.559	0.661
	MSA.OTS.For					8.26	15.379	0.552	0.352	0.765

Construct	Items	N	N of Items	Cronbach's Alpha	Cronbach's Alpha (Standardized Items)	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
T2										
Ideal Self										
Classical Ideal Self	Cl.IS.Plaus	172	4	.890	0.89	11.81	19.828	0.69	0.502	0.884
	Cl.IS.Vivid					12.13	18.767	0.772	0.598	0.854
	Cl.IS.Des					11.42	18.362	0.747	0.599	0.863
	Cl.IS.Freq					12.07	16.896	0.831	0.694	0.83
Dialect Ideal Self	D.IS.Vivid	169	4	.889	0.889	12.12	19.045	0.807	0.688	0.839
	D.IS.Freq					12.25	18.2	0.813	0.709	0.835
	D.IS.Plaus					12.12	21.01	0.665	0.452	0.89
	D.IS.Desire					11.66	19.453	0.747	0.57	0.861
MSA Ideal Self	MSA.IS.Vivid	166	4	.904	0.904	10.75	22.248	0.815	0.71	0.864
	MSA.IS.Freq					10.97	23.169	0.802	0.662	0.869
	MSA.IS.Plaus					10.76	23.966	0.781	0.66	0.877
	MSA.IS.Desire					10.28	23.474	0.739	0.579	0.892
Current Self										
Classical Current Self	Cl.CS.Ab	172	3	.850	0.851	6.48	7.631	0.776	0.652	0.739
	Cl.CS.Current					6.74	8.229	0.621	0.386	0.884
	Cl.CS.Conf					6.4	7.41	0.77	0.649	0.742
Dialect Current Self	D.CS.Ab	170	3	.877	0.88	6.55	8.178	0.817	0.684	0.783
	D.CS.Current					6.72	7.976	0.705	0.503	0.884
	D.CS.Conf					6.54	7.93	0.777	0.647	0.814
MSA Current Self	MSA.CS.Current	167	3	0.879	0.88	6.08	8.626	0.703	0.496	0.887
	MSA.CS.Conf					5.8	8.332	0.79	0.66	0.808
	MSA.CS.Ab					6.1	8.526	0.811	0.68	0.791
Ought-To Self										
Classical OT Self	Cl.OTS.Cons	171	4	0.713	0.712	9.26	13.945	0.516	0.271	0.643
	Cl.OTS.Diss					8.88	11.339	0.603	0.373	0.58
	Cl.OTS.Forc					9.41	14.373	0.406	0.179	0.703
	Cl.OTS.SocPress					7.4	13.323	0.483	0.27	0.66
Dialect OT Self	D.OTS.Force	170	4	0.784	0.786	8.99	15.734	0.545	0.362	0.754
	D.OTS.Diss					8.64	13.474	0.74	0.589	0.647
	D.OTS.SocPress					7.19	17.08	0.42	0.248	0.816
	D.OTS.Cons					9.28	15.364	0.685	0.563	0.688
MSA OT Self	MSA.OTS.SocPress	166	4	0.819	0.828	6.42	15.36	0.554	0.339	0.826
	MSA.OTS.Cons					7.93	17.129	0.681	0.52	0.761
	MSA.OTS.Diss					7.56	15.157	0.74	0.561	0.726
	MSA.OTS.For					7.82	16.658	0.628	0.432	0.779

C.9. Self-Rated Proficiency Items Descriptive Statistics

		N	Min.	Max.	Mean	SE	SD	Skew	SE	z-skew	Kurt	SE	z-kurt
T1	Reading Words	265	1	6	5.35	0.06	0.942	-1.65	0.15	-11	3.147	0.298	10.56
	MSA (Writing)	264	1	6	5.08	0.07	1.177	-1.36	0.15	-9.07	1.488	0.299	4.98
	MSA (Speaking)	264	1	6	4.47	0.08	1.305	-0.619	0.15	-4.13	-0.207	0.299	-0.69
	MSA (Reading Comprehension)	264	1	6	4.56	0.08	1.312	-0.794	0.15	-5.29	0.031	0.299	0.10
	MSA (Listening Comprehension)	265	1	6	4.62	0.08	1.271	-0.849	0.15	-5.66	0.427	0.298	1.43
	MSA v. Dialect	264	1	6	3.39	0.11	1.809	0.121	0.15	0.807	-1.329	0.299	-4.44
	Dialect (Speaking)	263	1	6	3.13	0.1	1.683	0.229	0.15	1.527	-1.085	0.299	-3.63
	Dialect (Listening Comprehension)	263	1	6	3.61	0.1	1.67	-0.098	0.15	-0.65	-1.151	0.299	-3.85
	Quran (Listening Comprehension)	264	1	6	3.72	0.1	1.621	-0.094	0.15	-0.63	-1.107	0.299	-3.70
	(Word Reading with Diacritics)	265	1	6	5.64	0.06	0.963	-2.995	0.15	-20	8.759	0.298	29.39
	Quran (Reading Comprehension)	265	1	6	3.6	0.1	1.607	-0.073	0.15	-0.49	-1.1	0.298	-3.69
T2	Reading Words	172	1	6	5.21	0.09	1.156	-1.797	0.185	-9.71	3.028	0.368	8.23
	MSA (Writing)	172	1	6	5.02	0.1	1.245	-1.524	0.185	-8.24	2.152	0.368	5.85
	MSA (Speaking)	171	1	6	4.61	0.11	1.448	-0.912	0.186	-4.9	-0.093	0.369	-0.25
	MSA (Reading Comprehension)	171	1	6	4.61	0.1	1.33	-0.797	0.186	-4.28	-0.141	0.369	-0.38
	MSA (Listening Comprehension)	170	1	6	4.59	0.1	1.33	-0.93	0.186	-5	0.396	0.37	1.07
	MSA v. Dialect	170	1	6	3.39	0.14	1.812	0.085	0.186	0.457	-1.347	0.37	-3.64
	Dialect (Speaking)	170	1	6	2.93	0.14	1.776	0.396	0.186	2.129	-1.17	0.37	-3.16
	Dialect (Listening Comprehension)	170	1	6	3.34	0.13	1.716	0.042	0.186	0.226	-1.258	0.37	-3.40
	Quran (Listening Comprehension)	170	1	6	3.72	0.13	1.657	-0.182	0.186	-0.98	-1.091	0.37	-2.95
	(Word Reading with Diacritics)	170	1	6	5.54	0.08	1.021	-2.541	0.186	-13.7	6.683	0.37	18.06
	Quran (Reading Comprehension)	170	1	6	3.51	0.13	1.697	0.053	0.186	0.285	-1.179	0.37	-3.19

C.10. Self-Rated Proficiency Correlation Matrices

Time 1	Reading Words	MSA (Writing)	MSA (Speaking)	MSA (Reading Comprehension)	MSA (Listening Comprehension)	MSA v. Dialect	Dialect (Speaking)	Dialect (Listening Comprehension)	Quran (Listening Comprehension)	(Word Reading with Diacritics)	Quran (Reading Comprehension)
Reading Words	1.000										
MSA (Writing)	.549**	1.000									
MSA (Speaking)	.416**	.665**	1.000								
MSA (Reading Comprehension)	.412**	.660**	.736**	1.000							
MSA (Listening Comprehension)	.402**	.579**	.643**	.745**	1.000						
MSA v. Dialect	.195**	.317**	.344**	.387**	.493**	1.000					
Dialect (Speaking)	0.053	.129*	.229**	.225**	.335**	.550**	1.000				
Dialect (Listening Comprehension)	0.115	.238**	.285**	.371**	.450**	.615**	.741**	1.000			
Quran (Listening Comprehension)	.361**	.360**	.380**	.440**	.464**	.469**	.324**	.382**	1.000		
(Word Reading with Diacritics)	.491**	.385**	.251**	.273**	.261**	0.087	0.001	0.088	.191**	1.000	
Quran (Reading Comprehension)	.293**	.322**	.349**	.432**	.449**	.410**	.321**	.337**	.796**	.170**	1.00

Significance indicators:

$p \leq .05 = *$

$p \leq .01 = **$

Time 2	Reading Words	MSA (Writing)	MSA (Speaking)	MSA (Reading Comprehension)	MSA (Listening Comprehension)	MSA v. Dialect	Dialect (Speaking)	Dialect (Listening Comprehension)	Quran (Listening Comprehension)	(Word Reading with Diacritics)	Quran (Reading Comprehension)
Reading Words	1										
MSA (Writing)	.629**	1									
MSA (Speaking)	.550**	.761**	1								
MSA (Reading Comprehension)	.535**	.748**	.730**	1							
MSA (Listening Comprehension)	.556**	.673**	.753**	.813**	1						
MSA v. Dialect	.268**	.340**	.467**	.410**	.470**	1					
Dialect (Speaking)	.153*	.168*	.282**	.269**	.301**	.523**	1				
Dialect (Listening Comprehension)	.188*	.205**	.302**	.296**	.330**	.583**	.802**	1			
Quran (Listening Comprehension)	.352**	.327**	.360**	.460**	.499**	.355**	.449**	.356**	1		
(Word Reading with Diacritics)	.534**	.419**	.426**	.424**	.496**	.172*	0.035	0.151	.180*	1	
Quran (Reading Comprehension)	.312**	.303**	.352**	.457**	.453**	.400**	.430**	.392**	.892**	.160*	1

Significance indicators:

$p \leq .05 = *$

$p \leq .01 = **$

		T2										
T1		Reading Words	MSA (Writing)	MSA (Speaking)	MSA (Reading Comprehension)	MSA (Listening Comprehension)	MSA v. Dialect	Dialect (Speaking)	Dialect (Listening Comprehension)	Quran (Listening Comprehension)	(Word Reading with Diacritics)	Quran (Reading Comprehension)
	Reading Words	.608**	.378**	.363**	.344**	.342**	.234**	0.043	0.12	.272**	.309**	.242**
	MSA (Writing)	.431**	.423**	.419**	.467**	.433**	.330**	0.085	.234**	.263**	.291**	.267**
	MSA (Speaking)	.423**	.464**	.558**	.492**	.532**	.535**	.154*	.245**	.344**	.345**	.357**
	MSA (Reading Comprehension)	.370**	.448**	.527**	.600**	.596**	.513**	.233**	.381**	.404**	.306**	.413**
	MSA (Listening Comprehension)	.447**	.543**	.596**	.630**	.604**	.453**	.182**	.262**	.436**	.346**	.408**
	MSA v. Dialect	.337**	.352**	.423**	.441**	.435**	.592**	.398**	.433**	.418**	.221**	.432**
	Dialect (Speaking)	.186**	.157*	.183**	.244**	.287**	.391**	.589**	.575**	.296**	0.116	.311**
	Dialect (Listening Comprehension)	.220**	.178*	.222**	.348**	.375**	.430**	.500**	.539**	.313**	.183**	.320**
	Quran (Listening Comprehension)	.273**	.209**	.228**	.328**	.319**	.368**	.278**	.282**	.674**	.134*	.704**
	(Word Reading with Diacritics)	.358**	.296**	.269**	.255**	.211**	.165*	0.035	0.094	.235**	.442**	.185**
	Quran (Reading Comprehension)	.246**	.262**	.232**	.357**	.298**	.319**	.263**	.242**	.633**	0.077	.666**

Significance indicators:

$p \leq .05 = *$

$p \leq .01 = **$

C.11. CFA Fit Indices – Self-Rated Proficiency

Fit	Modification	N	Model Fit (Robust)	df	p	CFI	TLI	RMSEA	SRMR
T1									
1	All items – 3 latent variables (MSA, Cl, D)		133.454	41	0.000	0.920	0.892	0.107	0.097
2	All items – 4 first order latent variables (MSA, Cl, D, Basic)	250	96.524	40	0.000	0.949	0.929	0.086	0.075
3	Basic items removed – 3 first order latent variables		54.555	24	0.000	0.949	0.924	0.093	0.085
3a	Basic items removed – 3 first order latent variables + MSAv.D crossloaded		54.195	23	0.000	0.97	0.953	0.081	0.046
T2									
2	All items – 4 first order latent variables (MSA, Cl, D, Basic)	152	114.424	40	0.000	0.929	0.902	0.118	0.100
3	Basic items removed – 3 first order latent variables		72.545	24	0.000	0.945	0.917	0.124	0.079
3a	Basic items removed – 3 first order latent variables + MSAv.D crossloaded		59.75	23	0.000	0.959	0.935	0.11	0.047

C.12. Reliability Scores – Self-Rated Proficiency Items

Construct	Items	N	N of Items	Cronbach's Alpha	Cronbach's Alpha based on Standardised Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
SR Prof All T1	Reading Words	250	11	0.867	0.873	41.7	93.189	0.462	0.438	0.863
	MSA (Writing)					41.95	87.029	0.644	0.616	0.852
	MSA (Speaking)					42.56	85.773	0.62	0.631	0.852
	MSA (Reading Comprehension)					42.47	84.66	0.677	0.669	0.848
	MSA (Listening Comprehension)					42.41	84.332	0.723	0.61	0.846
	MSA v. Dialect					43.7	79.567	0.605	0.496	0.854
	Dialect (Speaking)					43.95	84.66	0.487	0.563	0.863
	Dialect (Listening Comprehension)					43.46	82.282	0.576	0.623	0.855
	Quran (Listening Comprehension)					43.35	81.907	0.615	0.658	0.852
	(Word Reading with Diacritics)					41.4	95.173	0.323	0.276	0.869
	Quran (Reading Comprehension)					43.48	83.062	0.579	0.637	0.855
SR Prof All T2	Reading Words	170	11	0.897	0.905	41.25	116.282	0.625	0.641	0.889
	MSA (Writing)					41.44	113.905	0.668	0.784	0.886
	MSA (Speaking)					41.83	109.16	0.724	0.752	0.882
	MSA (Reading Comprehension)					41.85	110.32	0.757	0.783	0.881
	MSA (Listening Comprehension)					41.85	109.357	0.795	0.787	0.879
	MSA v. Dialect					43.05	107.956	0.585	0.451	0.892
	Dialect (Speaking)					43.52	109.375	0.558	0.691	0.893
	Dialect (Listening Comprehension)					43.11	109.532	0.578	0.69	0.892
	Quran (Listening Comprehension)					42.72	108.533	0.636	0.818	0.887
	(Word Reading with Diacritics)					40.91	120.779	0.509	0.568	0.894
	Quran (Reading Comprehension)					42.94	108.582	0.616	0.811	0.889

C.13. Descriptives Statistics – Proficiency Tests

Descriptives - YN Items T1 (N = 264)													
Item	Mean	SE	SD	Var.	Range	Min.	Max.	Skew.	Std Error	z-skew	Kurtosis	Std Error	z-kurt
1a	0.34	0.029	0.476	0.227	1 0	1	1	0.657	0.15	4.380	-1.580	0.299	-5.284
1b	0.25	0.027	0.432	0.186	1 0	1	1	1.185	0.15	7.900	-0.600	0.299	-2.007
1c	0.32	0.029	0.468	0.219	1 0	1	1	0.766	0.15	5.107	-1.423	0.299	-4.759
1e	0.77	0.026	0.422	0.178	1 0	1	1	-1.283	0.15	-8.553	-0.356	0.299	-1.191
1f	0.72	0.028	0.452	0.204	1 0	1	1	-0.963	0.15	-6.420	-1.081	0.299	-3.615
1h	0.25	0.027	0.436	0.19	1 0	1	1	1.138	0.15	7.587	-0.710	0.299	-2.375
1i	0.21	0.025	0.41	0.168	1 0	1	1	1.416	0.15	9.440	0.006	0.299	0.020
1j	0.32	0.029	0.468	0.219	1 0	1	1	0.766	0.15	5.107	-1.423	0.299	-4.759
1l	0.15	0.022	0.356	0.126	1 0	1	1	1.997	0.15	13.313	2.003	0.299	6.699
1m	0.24	0.026	0.429	0.184	1 0	1	1	1.209	0.15	8.060	-0.543	0.299	-1.816
1n	0.42	0.03	0.495	0.245	1 0	1	1	0.308	0.15	2.053	-1.920	0.299	-6.421
1p	0.31	0.029	0.465	0.216	1 0	1	1	0.804	0.15	5.360	-1.364	0.299	-4.562
2b	0.38	0.03	0.487	0.237	1 0	1	1	0.486	0.15	3.240	-1.777	0.299	-5.943
2d	0.14	0.021	0.344	0.118	1 0	1	1	2.131	0.15	14.207	2.562	0.299	8.569
2e	0.23	0.026	0.422	0.178	1 0	1	1	1.283	0.15	8.553	-0.356	0.299	-1.191
2f	0.62	0.03	0.485	0.235	1 0	1	1	-0.519	0.15	-3.460	-1.744	0.299	-5.833
2g	0.23	0.026	0.425	0.18	1 0	1	1	1.258	0.15	8.387	-0.420	0.299	-1.405
2h	0.19	0.024	0.393	0.154	1 0	1	1	1.595	0.15	10.633	0.547	0.299	1.829
2i	0.33	0.029	0.47	0.22	1 0	1	1	0.748	0.15	4.987	-1.452	0.299	-4.856
2l	0.58	0.03	0.495	0.245	1 0	1	1	-0.324	0.15	-2.160	-1.909	0.299	-6.385
2n	0.74	0.027	0.438	0.192	1 0	1	1	-1.115	0.15	-7.433	-0.762	0.299	-2.548
2o	0.32	0.029	0.467	0.218	1 0	1	1	0.785	0.15	5.233	-1.394	0.299	-4.662
2p	0.37	0.03	0.483	0.233	1 0	1	1	0.553	0.15	3.687	-1.707	0.299	-5.709
2q	0.31	0.029	0.464	0.215	1 0	1	1	0.823	0.15	5.487	-1.332	0.299	-4.455
3a	0.33	0.029	0.471	0.222	1 0	1	1	0.729	0.15	4.860	-1.479	0.299	-4.946
3c	0.71	0.028	0.455	0.207	1 0	1	1	-0.922	0.15	-6.147	-1.159	0.299	-3.876
3d	0.25	0.027	0.432	0.186	1 0	1	1	1.185	0.15	7.900	-0.600	0.299	-2.007
3e	0.29	0.028	0.454	0.206	1 0	1	1	0.942	0.15	6.280	-1.121	0.299	-3.749
3f	0.51	0.031	0.501	0.251	1 0	1	1	-0.03	0.15	-0.200	-2.014	0.299	-6.736
3i	0.32	0.029	0.467	0.218	1 0	1	1	0.785	0.15	5.233	-1.394	0.299	-4.662
3j	0.55	0.031	0.499	0.249	1 0	1	1	-0.199	0.15	-1.327	-1.975	0.299	-6.605
3k	0.87	0.021	0.336	0.113	1 0	1	1	-2.229	0.15	-14.860	2.992	0.299	10.01
3m	0.35	0.029	0.477	0.228	1 0	1	1	0.64	0.15	4.267	-1.603	0.299	-5.361
3p	0.67	0.029	0.472	0.223	1 0	1	1	-0.711	0.15	-4.740	-1.506	0.299	-5.037
3q	0.81	0.024	0.393	0.154	1 0	1	1	-1.595	0.15	-10.633	0.547	0.299	1.829
4a	0.31	0.029	0.465	0.216	1 0	1	1	0.804	0.15	5.360	-1.364	0.299	-4.562
4b	0.29	0.028	0.455	0.207	1 0	1	1	0.922	0.15	6.147	-1.159	0.299	-3.876
4d	0.44	0.031	0.497	0.247	1 0	1	1	0.246	0.15	1.640	-1.955	0.299	-6.538
4e	0.13	0.021	0.336	0.113	1 0	1	1	2.229	0.15	14.860	2.992	0.299	10.01
4g	0.91	0.018	0.293	0.086	1 0	1	1	-2.784	0.15	-18.560	5.796	0.299	19.39

4h	0.12	0.02	0.323	0.104	1	0	1	2.39	0.15	15.933	3.742	0.299	12.52
4i	0.19	0.024	0.393	0.154	1	0	1	1.595	0.15	10.633	0.547	0.299	1.829
4j	0.34	0.029	0.476	0.227	1	0	1	0.657	0.15	4.380	-1.580	0.299	-5.284
4m	0.29	0.028	0.455	0.207	1	0	1	0.922	0.15	6.147	-1.159	0.299	-3.876
4n	0.23	0.026	0.425	0.18	1	0	1	1.258	0.15	8.387	-0.420	0.299	-1.405
4o	0.34	0.029	0.474	0.224	1	0	1	0.693	0.15	4.620	-1.531	0.299	-5.120
4p	0.43	0.031	0.496	0.246	1	0	1	0.293	0.15	1.953	-1.929	0.299	-6.452
4q	0.55	0.031	0.499	0.249	1	0	1	-0.199	0.15	-1.327	-1.975	0.299	-6.605
5b	0.87	0.021	0.336	0.113	1	0	1	-2.229	0.15	-14.860	2.992	0.299	10.01
5c	0.38	0.03	0.486	0.236	1	0	1	0.503	0.15	3.353	-1.761	0.299	-5.890
5d	0.38	0.03	0.486	0.236	1	0	1	0.503	0.15	3.353	-1.761	0.299	-5.890
5g	0.29	0.028	0.455	0.207	1	0	1	0.922	0.15	6.147	-1.159	0.299	-3.876
5h	0.58	0.03	0.495	0.245	1	0	1	-0.308	0.15	-2.053	-1.920	0.299	-6.421
5j	0.21	0.025	0.407	0.166	1	0	1	1.445	0.15	9.633	0.087	0.299	0.291
5l	0.27	0.027	0.446	0.199	1	0	1	1.026	0.15	6.840	-0.954	0.299	-3.191
5m	0.43	0.031	0.496	0.246	1	0	1	0.277	0.15	1.847	-1.938	0.299	-6.482
5n	0.58	0.03	0.495	0.245	1	0	1	-0.324	0.15	-2.160	-1.909	0.299	-6.385
5o	0.32	0.029	0.468	0.219	1	0	1	0.766	0.15	5.107	-1.423	0.299	-4.759
5p	0.2	0.025	0.401	0.161	1	0	1	1.503	0.15	10.020	0.260	0.299	0.870
5q	0.6	0.03	0.491	0.241	1	0	1	-0.404	0.15	-2.693	-1.851	0.299	-6.191
Note: Pseudoword items are removed and real word item numbers are colour coded according to word frequency band.													

Descriptives - YN Items T2 (N=172)													
Item	Mean	SE	SD	Var.	Range	Min.	Max.	Skew.	Std Error	p-value	Kurtosis	Std Error	p-value
1a	0.4	0.037	0.49	0.24	1 0	1	1	0.432	0.185	2.335	-1.835	0.368	-4.986
1b	0.32	0.036	0.468	0.219	1 0	1	1	0.78	0.185	4.216	-1.409	0.368	-3.829
1c	0.3	0.035	0.458	0.21	1 0	1	1	0.899	0.185	4.859	-1.206	0.368	-3.277
1e	0.84	0.028	0.365	0.133	1 0	1	1	-1.903	0.185	-10.286	1.639	0.368	4.454
1f	0.74	0.034	0.441	0.194	1 0	1	1	-1.094	0.185	-5.914	-0.812	0.368	-2.207
1h	0.26	0.034	0.441	0.194	1 0	1	1	1.094	0.185	5.914	-0.812	0.368	-2.207
1i	0.2	0.031	0.404	0.163	1 0	1	1	1.486	0.185	8.032	0.211	0.368	0.573
1j	0.38	0.037	0.486	0.236	1 0	1	1	0.508	0.185	2.746	-1.763	0.368	-4.791
1l	0.18	0.029	0.386	0.149	1 0	1	1	1.678	0.185	9.070	0.827	0.368	2.247
1m	0.24	0.033	0.431	0.186	1 0	1	1	1.201	0.185	6.492	-0.563	0.368	-1.530
1n	0.39	0.037	0.489	0.239	1 0	1	1	0.457	0.185	2.470	-1.812	0.368	-4.924
1p	0.3	0.035	0.458	0.21	1 0	1	1	0.899	0.185	4.859	-1.206	0.368	-3.277
2b	0.38	0.037	0.486	0.236	1 0	1	1	0.508	0.185	2.746	-1.763	0.368	-4.791
2d	0.17	0.029	0.375	0.141	1 0	1	1	1.786	0.185	9.654	1.203	0.368	3.269
2e	0.22	0.032	0.416	0.173	1 0	1	1	1.357	0.185	7.335	-0.16	0.368	-0.435
2f	0.66	0.036	0.476	0.227	1 0	1	1	-0.667	0.185	-3.605	-1.573	0.368	-4.274
2g	0.31	0.035	0.465	0.217	1 0	1	1	0.809	0.185	4.373	-1.362	0.368	-3.701
2h	0.21	0.031	0.408	0.166	1 0	1	1	1.442	0.185	7.795	0.079	0.368	0.215
2i	0.38	0.037	0.486	0.236	1 0	1	1	0.508	0.185	2.746	-1.763	0.368	-4.791
2l	0.52	0.038	0.501	0.251	1 0	1	1	-0.07	0.185	-0.378	-2.019	0.368	-5.486
2n	0.71	0.035	0.455	0.207	1 0	1	1	-0.93	0.185	-5.027	-1.149	0.368	-3.122
2o	0.32	0.036	0.468	0.219	1 0	1	1	0.78	0.185	4.216	-1.409	0.368	-3.829
2p	0.31	0.035	0.463	0.214	1 0	1	1	0.838	0.185	4.530	-1.313	0.368	-3.568
2q	0.39	0.037	0.489	0.239	1 0	1	1	0.457	0.185	2.470	-1.812	0.368	-4.924
3a	0.4	0.037	0.492	0.242	1 0	1	1	0.407	0.185	2.200	-1.856	0.368	-5.043
3c	0.77	0.032	0.424	0.18	1 0	1	1	-1.277	0.185	-6.903	-0.373	0.368	-1.014
3d	0.31	0.035	0.465	0.217	1 0	1	1	0.809	0.185	4.373	-1.362	0.368	-3.701
3e	0.38	0.037	0.486	0.236	1 0	1	1	0.508	0.185	2.746	-1.763	0.368	-4.791
3f	0.49	0.038	0.501	0.251	1 0	1	1	0.023	0.185	0.124	-2.023	0.368	-5.497
3i	0.37	0.037	0.483	0.233	1 0	1	1	0.56	0.185	3.027	-1.706	0.368	-4.636
3j	0.61	0.037	0.489	0.239	1 0	1	1	-0.457	0.185	-2.470	-1.812	0.368	-4.924
3k	0.8	0.031	0.404	0.163	1 0	1	1	-1.486	0.185	-8.032	0.211	0.368	0.573
3m	0.32	0.036	0.468	0.219	1 0	1	1	0.78	0.185	4.216	-1.409	0.368	-3.829
3p	0.65	0.037	0.48	0.23	1 0	1	1	-0.613	0.185	-3.314	-1.643	0.368	-4.465
3q	0.84	0.028	0.37	0.137	1 0	1	1	-1.843	0.185	-9.962	1.413	0.368	3.840
4a	0.34	0.036	0.474	0.225	1 0	1	1	0.695	0.185	3.757	-1.535	0.368	-4.171
4b	0.48	0.038	0.501	0.251	1 0	1	1	0.07	0.185	0.378	-2.019	0.368	-5.486
4d	0.42	0.038	0.496	0.246	1 0	1	1	0.309	0.185	1.670	-1.927	0.368	-5.236
4e	0.16	0.028	0.365	0.133	1 0	1	1	1.903	0.185	10.286	1.639	0.368	4.454
4g	0.87	0.026	0.341	0.117	1 0	1	1	-2.171	0.185	-11.735	2.746	0.368	7.462
4h	0.16	0.028	0.365	0.133	1 0	1	1	1.903	0.185	10.286	1.639	0.368	4.454
4i	0.15	0.027	0.353	0.125	1 0	1	1	2.03	0.185	10.973	2.147	0.368	5.834

4j	0.5	0.038	0.501	0.251	1	0	1	0	0.185	0.000	-2.024	0.368	-5.500
4m	0.25	0.033	0.434	0.189	1	0	1	1.165	0.185	6.297	-0.651	0.368	-1.769
4n	0.26	0.033	0.438	0.191	1	0	1	1.129	0.185	6.103	-0.734	0.368	-1.995
4o	0.58	0.038	0.496	0.246	1	0	1	-0.309	0.185	-1.670	-1.927	0.368	-5.236
4p	0.4	0.037	0.492	0.242	1	0	1	0.407	0.185	2.200	-1.856	0.368	-5.043
4q	0.56	0.038	0.497	0.247	1	0	1	-0.26	0.185	-1.405	-1.955	0.368	-5.313
5b	0.88	0.025	0.328	0.108	1	0	1	-2.329	0.185	-12.589	3.464	0.368	9.413
5c	0.53	0.038	0.5	0.25	1	0	1	-0.141	0.185	-0.762	-2.004	0.368	-5.446
5d	0.41	0.038	0.494	0.244	1	0	1	0.357	0.185	1.930	-1.894	0.368	-5.147
5g	0.23	0.032	0.424	0.18	1	0	1	1.277	0.185	6.903	-0.373	0.368	-1.014
5h	0.62	0.037	0.486	0.236	1	0	1	-0.508	0.185	-2.746	-1.763	0.368	-4.791
5j	0.16	0.028	0.37	0.137	1	0	1	1.843	0.185	9.962	1.413	0.368	3.840
5l	0.34	0.036	0.476	0.227	1	0	1	0.667	0.185	3.605	-1.573	0.368	-4.274
5m	0.41	0.038	0.493	0.243	1	0	1	0.382	0.185	2.065	-1.876	0.368	-5.098
5n	0.61	0.037	0.489	0.239	1	0	1	-0.457	0.185	-2.470	-1.812	0.368	-4.924
5o	0.34	0.036	0.476	0.227	1	0	1	0.667	0.185	3.605	-1.573	0.368	-4.274
5p	0.22	0.032	0.416	0.173	1	0	1	1.357	0.185	7.335	-0.16	0.368	-0.435
5q	0.58	0.038	0.495	0.245	1	0	1	-0.333	0.185	-1.800	-1.912	0.368	-5.196

Note: Pseudoword items are removed and real word item numbers are colour coded according to word frequency band.

Descriptives - Cloze Tests															
Activity	Item	N	Mean	SE	SD	Var.	Range	Min.	Max.	Skew.	Std Error	z-value	Kurtosis	Std Error	z-value
1	1a	263	0.74	0.027	0.439	0.192	1	0	1	-1.109	0.15	-7.393	-0.776	0.299	-2.595
	1b		0.65	0.029	0.477	0.227				-0.651	0.15	-4.340	-1.588	0.299	-5.311
	1c		0.56	0.031	0.497	0.247				-0.239	0.15	-1.593	-1.958	0.299	-6.548
	1d		0.6	0.03	0.491	0.242				-0.398	0.15	-2.653	-1.856	0.299	-6.207
	1e		0.72	0.028	0.451	0.203				-0.978	0.15	-6.520	-1.052	0.299	-3.518
	1f		0.55	0.031	0.499	0.249				-0.192	0.15	-1.280	-1.978	0.299	-6.615
	1g		0.53	0.031	0.5	0.25				-0.130	0.15	-0.867	-1.998	0.299	-6.682
	1h		0.78	0.026	0.415	0.173				-1.356	0.15	-9.040	-0.163	0.299	-0.545
2	2a	263	0.79	0.025	0.407	0.166				-1.439	0.15	-9.593	0.070	0.299	0.234
	2b		0.47	0.031	0.5	0.25				0.115	0.15	0.767	-2.002	0.299	-6.696
	2c		0.4	0.03	0.49	0.24				0.430	0.15	2.867	-1.829	0.299	-6.117
	2d		0.73	0.028	0.447	0.2				-1.021	0.15	-6.807	-0.966	0.299	-3.231
	2e		0.48	0.031	0.501	0.251				0.084	0.15	0.560	-2.008	0.299	-6.716
3	3a	263	0.51	0.031	0.501	0.251				-0.054	0.15	-0.360	-2.012	0.299	-6.729
	3b		0.59	0.03	0.494	0.244				-0.349	0.15	-2.327	-1.892	0.299	-6.328
	3c		0.52	0.031	0.501	0.251				-0.084	0.15	-0.560	-2.008	0.299	-6.716
	3d		0.51	0.031	0.501	0.251				-0.023	0.15	-0.153	-2.015	0.299	-6.739
	3e		0.51	0.031	0.501	0.251				-0.038	0.15	-0.253	-2.014	0.299	-6.736
	3f		0.52	0.031	0.501	0.251				-0.084	0.15	-0.560	-2.008	0.299	-6.716
	3g		0.25	0.027	0.437	0.191				1.132	0.15	7.547	-0.724	0.299	-2.421
	3h		0.3	0.028	0.458	0.209				0.896	0.15	5.973	-1.207	0.299	-4.037
	3i		0.55	0.031	0.498	0.248				-0.208	0.15	-1.387	-1.972	0.299	-6.595
	3j		0.52	0.031	0.5	0.25				-0.100	0.15	-0.667	-2.005	0.299	-6.706
	3k		0.43	0.031	0.495	0.245				0.302	0.15	2.013	-1.924	0.299	-6.435
	3l		0.59	0.03	0.492	0.242				-0.381	0.15	-2.540	-1.869	0.299	-6.251
	1Total	265	5.09	0.137	2.233	4.984	8	0	8	-0.446	0.15	-2.973	-0.633	0.298	-2.124
	2Total		2.84	0.091	1.484	2.202	5	0	5	-0.110	0.15	-0.733	-0.889	0.298	-2.983
	3Total		5.76	0.214	3.492	12.191	12	0	12	0.183	0.15	1.220	-0.936	0.298	-3.141

C.14. Correlations between Proficiency Measures

Table 4.6. Correlations between Proficiency Measures

Spearman's rho		Y/N Test T1	Cloze Test T1	SR Prof MSA T1	SR Prof Dialect T1	SR Prof Classical T1	Y/N Test T2	Cloze Test T2	SR Prof MSA T2	SR Prof Dialect T2	SR Prof Classical T2
Cloze T1	Corr. Coeff.	.598**									
	N	264									
SR Prof MSA T1	Corr. Coeff.	.632**	.483**								
	N	249	249								
SR Prof D T1	Corr. Coeff.	.515**	.261**	.479**							
	N	249	249	250							
SR Prof Cl T1	Corr. Coeff.	.627**	.407**	.613**	.477**						
	N	249	249	250	250						
SR Prof Overall T1	Corr. Coeff.	.720**	.481**	.872**	.687**	.869**					
	N	249	249	250	250	250					
Cloze T2	Corr. Coeff.	.490**	.707**	.402**	.206**	.393**	.531**				
	N	167	167	159	159	159	171				
SR Prof MSA T2	Corr. Coeff.	.534**	.410**	.704**	.360**	.417**	.443**	.397**			
	N	152	152	149	149	149	151	150			
SR Prof D T2	Corr. Coeff.	.478**	.162*	.362**	.566**	.353**	.394**	.200*	.355**		
	N	152	152	149	149	149	151	150	152		
SR Prof Cl T2	Corr. Coeff.	.527**	.336**	.498**	.396**	.731**	.449**	.339**	.550**	.527**	
	N	152	152	149	149	149	151	150	152	152	
SR Prof Overall T2	Corr. Coeff.	.601**	.373**	.585**	.484**	.699**	.503**	.374**	.687**	.659**	.961**
	N	152	152	149	149	149	151	150	152	152	152

Significance indicators:

$p \leq .05 = *$

$p \leq .01 = **$

C.15. Descriptive Statistics – Motivation Scale Average-Scores

	Average scores (ex. Plaus)	N	Min	Max	M	SE	SD	Skew	SE Skew	z(skew)	Kurt	SE Kurt	z(kurt)
T1	CI CS	264	1.00	6.00	3.53	0.08	1.24	-0.09	0.15	-0.57	-0.65	0.30	-2.17
	CI IS	263	1.00	6.00	4.19	0.07	1.21	-0.45	0.15	-2.99	-0.54	0.30	-1.80
	CI OTS	264	1.00	6.00	3.06	0.07	1.14	0.33	0.15	2.18	-0.35	0.30	-1.16
	D CS	263	1.00	6.00	3.34	0.08	1.30	0.14	0.15	0.96	-0.72	0.30	-2.40
	D IS	265	1.00	6.00	4.32	0.08	1.33	-0.51	0.15	-3.37	-0.57	0.30	-1.93
	D OTS	263	1.00	6.00	2.87	0.07	1.19	0.54	0.15	3.58	-0.14	0.30	-0.46
	MSA CS	261	1.00	6.00	3.00	0.08	1.36	0.33	0.15	2.21	-0.58	0.30	-1.92
	MSA IS	261	1.00	6.00	3.89	0.09	1.49	-0.33	0.15	-2.19	-0.93	0.30	-3.09
	MSA OTS	260	1.00	6.00	2.54	0.07	1.19	0.56	0.15	3.68	-0.34	0.30	-1.12
T2	CI CS	173	1.00	6.00	3.25	0.10	1.34	0.08	0.19	0.42	-0.85	0.37	-2.33
	CI IS	174	1.00	6.00	3.94	0.11	1.49	-0.48	0.18	-2.62	-0.71	0.37	-1.94
	CI OTS	174	1.00	6.00	2.89	0.09	1.16	0.32	0.18	1.73	-0.41	0.37	-1.12
	D CS	172	1.00	6.00	3.29	0.10	1.38	0.06	0.19	0.32	-0.69	0.37	-1.89
	D IS	172	1.00	6.00	4.02	0.12	1.53	-0.45	0.19	-2.41	-0.73	0.37	-1.99
	D OTS	172	1.00	6.00	2.84	0.10	1.26	0.64	0.19	3.48	0.05	0.37	0.14
	MSA CS	167	1.00	6.00	3.00	0.11	1.42	0.30	0.19	1.57	-0.67	0.37	-1.78
	MSA IS	167	1.00	6.00	3.58	0.13	1.63	-0.17	0.19	-0.92	-1.13	0.37	-3.03
	MSA OTS	166	1.00	6.00	2.48	0.10	1.30	0.83	0.19	4.42	0.16	0.38	0.42

C.16. Pairwise T-Test Results – Comparison of Self-Concept Average-Scores – Between Varieties

			Mean	N	Std. Deviation	Std. Error Mean	Mean Diff	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	Corr.	Sig.	Effect Size
											Lower	Upper					
T1	Pair 1	Cl CS	3.53	263	1.25	0.08	0.18	0.89	0.05	0.08	0.29	3.36	262	0.001***	0.758	0.000	.21
		D CS	3.34	263	1.30	0.08											
	Pair 2	D CS	3.34	260	1.30	0.08	0.33	0.84	0.05	0.23	0.44	6.39	259	0.000***	0.798	0.000	.40
		MSA CS	3	260	1.35	0.08											
	Pair 3	MSA CS	3	261	1.36	0.08	-0.52	0.93	0.06	-0.63	-0.41	-8.99	260	0.000***	0.745	0.000	.56
		Cl CS	3.52	261	1.24	0.08											
	Pair 1	Cl IS	4.19	263	1.21	0.07	-0.15	0.80	0.05	-0.25	-0.05	-3.03	262	0.003**	0.802	0.000	.19
		D IS	4.34	263	1.31	0.08											
	Pair 2	D IS	4.33	261	1.32	0.08	0.44	1.01	0.06	0.31	0.56	6.98	260	0.000***	0.746	0.000	.43
		MSA IS	3.89	261	1.49	0.09											
	Pair 3	MSA IS	3.9	260	1.49	0.09	-0.28	1.04	0.06	-0.41	-0.16	-4.42	259	0.000***	0.722	0.000	.28
		Cl IS	4.19	260	1.21	0.07											
	Pair 1	Cl OTS	3.07	263	1.14	0.07	0.20	0.75	0.05	0.11	0.29	4.25	262	0.000***	0.790	0.000	.26
		D OTS	2.87	263	1.19	0.07											
	Pair 2	D OTS	2.88	260	1.19	0.07	0.34	0.78	0.05	0.25	0.44	7.09	259	0.000***	0.783	0.000	.43
		MSA OTS	2.54	260	1.19	0.07											
	Pair 3	MSA OTS	2.54	260	1.19	0.07	-0.53	0.90	0.06	-0.64	-0.42	-9.47	259	0.000***	0.699	0.000	.59
		Cl OTS	3.07	260	1.13	0.07											

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

			Mean	N	Std. Deviation	Std. Error Mean	Mean Diff	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	Corr.	Sig.	Effect Size
											Lower	Upper					
T2	Pair 1	CI CS	3.24	171	1.34	0.10	-0.04	0.97	0.07	-0.19	0.11	-0.55	170	0.581	0.747	0.000	-
		D CS	3.28	171	1.37	0.11											
	Pair 2	D CS	3.26	167	1.38	0.11	0.27	0.92	0.07	0.13	0.41	3.75	166	0.000***	0.783	0.000	.28
		MSA CS	3	167	1.42	0.11											
	Pair 3	MSA CS	3	166	1.42	0.11	-0.23	0.95	0.07	-0.38	-0.08	-3.12	165	0.002**	0.764	0.000	.24
		CI CS	3.23	166	1.35	0.10											
	Pair 1	CI IS	3.94	172	1.49	0.11	-0.08	0.98	0.07	-0.23	0.07	-1.06	171	0.290	0.789	0.000	-
		D IS	4.02	172	1.53	0.12											
	Pair 2	D IS	4.01	167	1.52	0.12	0.43	1.18	0.09	0.25	0.61	4.66	166	0.000***	0.719	0.000	.36
		MSA IS	3.58	167	1.63	0.13											
	Pair 3	MSA IS	3.58	167	1.63	0.13	-0.37	1.11	0.09	-0.54	-0.20	-4.24	166	0.000***	0.748	0.000	.33
		CI IS	3.95	167	1.49	0.12											
	Pair 1	CI OTS	2.9	171	1.16	0.09	0.06	0.81	0.06	-0.06	0.18	0.92	170	0.357	0.779	0.000	-
		D OTS	2.85	171	1.25	0.10											
	Pair 2	D OTS	2.85	166	1.26	0.10	0.37	0.94	0.07	0.23	0.51	5.06	165	0.000***	0.728	0.000	.39
		MSA OTS	2.48	166	1.3	0.10											
	Pair 3	MSA OTS	2.48	166	1.3	0.10	-0.45	1.05	0.08	-0.61	-0.29	-5.49	165	0.000***	0.638	0.000	.43
		CI OTS	2.93	166	1.16	0.09											

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

C.17. Self-Concept Average-Score Descriptive Statistics – Gender Differences

Gender	Variable	N	Min.	Max.	Mean	Std. Error	Std. Deviation	Skewness	Std. Error	z-skew	Kurtosis	Std. Error	z-kurt
Males	CI CS	86	0.33	6.00	3.616	0.137	1.270	-0.299	0.26	-1.150	-0.366	0.514	-0.712
	CI IS	84	1.00	6.00	4.341	0.133	1.215	-0.566	0.263	-2.152	-0.364	0.52	-0.700
	CI OTS	85	1.00	6.00	3.212	0.129	1.191	0.299	0.261	1.146	-0.497	0.517	-0.961
	D CS	84	1.00	6.00	3.476	0.148	1.355	-0.028	0.263	-0.106	-0.805	0.52	-1.548
	D IS	86	1.00	6.00	4.484	0.137	1.275	-0.503	0.26	-1.935	-0.672	0.514	-1.307
	D OTS	84	1.00	6.00	2.991	0.127	1.165	0.402	0.263	1.529	-0.474	0.52	-0.912
	MSA CS	82	1.00	6.00	3.061	0.149	1.348	0.091	0.266	0.342	-0.63	0.526	-1.198
	MSA IS	82	1.00	6.00	4.163	0.159	1.438	-0.531	0.266	-1.996	-0.542	0.526	-1.030
	MSA OTS	81	1.00	5.25	2.769	0.136	1.225	0.086	0.267	0.322	-1.112	0.529	-2.102
	CI CS T2	58	1.00	6.00	3.454	0.174	1.326	-0.03	0.314	-0.096	-0.796	0.618	-1.288
	CI IS T2	58	1.00	6.00	4.253	0.190	1.445	-0.81	0.314	-2.580	-0.161	0.618	-0.261
	CI OTS T2	59	0.75	6.00	3.081	0.165	1.270	0.053	0.311	0.170	-0.574	0.613	-0.936
	D CS T2	57	1.00	6.00	3.661	0.182	1.374	-0.195	0.316	-0.617	-0.486	0.623	-0.780
	D IS T2	57	1.00	6.00	4.304	0.192	1.449	-0.64	0.316	-2.025	-0.373	0.623	-0.599
	D OTS T2	57	1.00	6.00	3.110	0.168	1.270	0.54	0.316	1.709	0.153	0.623	0.246
	MSA CS T2	55	1.00	6.00	3.091	0.199	1.475	0.373	0.322	1.158	-0.624	0.634	-0.984
	MSA IS T2	55	1.00	6.00	3.812	0.225	1.668	-0.309	0.322	-0.960	-1.229	0.634	-1.938
	MSA OTS T2	55	0.75	6.00	2.709	0.180	1.332	0.757	0.322	2.351	0.077	0.634	0.121

Gender	Variable	N	Min.	Max.	Mean	Std. Error	Std. Deviation	Skewness	Std. Error	z-skew	Kurtosis	Std. Error	z-kurt
Females	CI CS	179	1.00	6.00	3.466	0.094	1.252	-0.03	0.182	-0.165	-0.658	0.361	-1.823
	CI IS	179	1.33	6.00	4.121	0.090	1.207	-0.405	0.182	-2.225	-0.573	0.361	-1.587
	CI OTS	179	1.00	6.00	2.993	0.083	1.113	0.323	0.182	1.775	-0.269	0.361	-0.745
	D CS	179	1.00	6.00	3.279	0.095	1.266	0.222	0.182	1.220	-0.625	0.361	-1.731
	D IS	179	1.00	6.00	4.246	0.101	1.346	-0.498	0.182	-2.736	-0.559	0.361	-1.548
	D OTS	179	1.00	6.00	2.817	0.090	1.198	0.612	0.182	3.363	0.066	0.361	0.183
	MSA CS	179	1.00	6.00	2.966	0.102	1.364	0.444	0.182	2.440	-0.508	0.361	-1.407
	MSA IS	179	1.00	6.00	3.769	0.112	1.502	-0.246	0.182	-1.352	-1.032	0.361	-2.859
	MSA OTS	179	1.00	6.00	2.436	0.087	1.161	0.796	0.182	4.374	0.325	0.361	0.900
	CI CS T2	116	1.00	6.00	3.161	0.125	1.341	0.124	0.225	0.551	-0.84	0.446	-1.883
	CI IS T2	116	1.00	6.00	3.782	0.139	1.492	-0.348	0.225	-1.547	-0.806	0.446	-1.807
	CI OTS T2	116	1.00	5.75	2.780	0.103	1.109	0.423	0.225	1.880	-0.225	0.446	-0.504
	D CS T2	115	1.00	6.00	3.113	0.126	1.347	0.175	0.226	0.774	-0.65	0.447	-1.454
	D IS T2	115	1.00	6.00	3.878	0.145	1.551	-0.354	0.226	-1.566	-0.826	0.447	-1.848
	D OTS T2	115	1.00	6.00	2.700	0.115	1.235	0.722	0.226	3.195	0.121	0.447	0.271
	MSA CS T2	112	1.00	6.00	2.949	0.131	1.389	0.244	0.228	1.070	-0.703	0.453	-1.552
	MSA IS T2	112	1.00	6.00	3.470	0.151	1.603	-0.122	0.228	-0.535	-1.055	0.453	-2.329
	MSA OTS T2	112	1.00	6.00	2.348	0.120	1.274	0.89	0.228	3.904	0.281	0.453	0.620

C.18. T-Test Results – Between Group Comparisons of Self-Concept scores – Gender Background

									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
T1	CI CS	0.159	0.691	1.007	170.278	0.316	0.164	0.163	-0.158	0.487
	CI IS	0.172	0.678	1.308	164.02	0.193	0.211	0.161	-0.108	0.530
	CI OTS	0.556	0.457	1.342	157.776	0.181	0.207	0.155	-0.098	0.513
	D CS	0.762	0.384	0.92	154.25	0.359	0.162	0.176	-0.186	0.510
	D IS	0.229	0.633	1.289	176.835	0.199	0.220	0.171	-0.117	0.557
	D OTS	0	0.988	1.036	168.746	0.302	0.162	0.156	-0.147	0.471
	MSA CS	0.005	0.942	0.357	161.033	0.721	0.065	0.181	-0.293	0.423
	MSA IS	0.839	0.361	1.937	165.824	0.054	0.379	0.196	-0.007	0.766
	MSA OTS	1.469	0.227	1.938	149.548	0.055	0.315	0.162	-0.006	0.635
T2	CI CS T2	0.04	0.841	1.287	114.072	0.201	0.281	0.218	-0.152	0.713
	CI IS T2	0.096	0.757	1.878	118.368	0.063	0.443	0.236	-0.024	0.911
	CI OTS T2	0.636	0.426	1.734	105.398	0.086	0.337	0.194	-0.048	0.723
	D CS T2	0.001	0.979	2.607	111.165	0.010	0.580	0.222	0.139	1.020
	D IS T2	0.254	0.615	1.854	119.999	0.066	0.448	0.242	-0.031	0.927
	D OTS T2	0.062	0.804	1.978	111.334	0.050	0.407	0.206	-0.001	0.815
	MSA CS T2	0.189	0.664	0.584	102.917	0.560	0.140	0.239	-0.335	0.614
	MSA IS T2	1	0.319	1.099	103.924	0.274	0.298	0.272	-0.240	0.837
	MSA OTS T2	0.048	0.828	1.766	102.7	0.080	0.383	0.217	-0.047	0.813

C.19. T-Test Results – Gender Split Comparison of Motivation Scale Average-Score Means

Paired Samples Test – Gender Split File (0 = Male; 1 = Female)															
Gender				Mean	SD	SE	95% CI		t	df	Sig. (2-tailed)	Corr.	Corr Sig.	Effect Size	
							Lower	Upper							
T1	0	Pair 1	CS	Cl - DS	0.18	0.85	0.09	-0.01	0.36	1.923	83	0.058	0.788	0.000	
		Pair 2		D - MSA	0.38	0.99	0.11	0.16	0.60	3.463	80	0.001***	0.728	0.000	0.38
		Pair 3		MSA - Cl	-0.57	0.90	0.10	-0.76	-0.37	-5.67	81	0.000***	0.756	0.000	-0.63
	1	Pair 1	CS	Cl - DS	0.19	0.90	0.07	0.05	0.32	2.755	178	0.006**	0.742	0.000	0.21
		Pair 2		D - MSA	0.31	0.77	0.06	0.20	0.43	5.446	178	0.000***	0.832	0.000	0.40
		Pair 3		MSA - Cl	-0.50	0.95	0.07	-0.64	-0.36	-7.03	178	0.000***	0.74	0.000	-0.53
	0	Pair 1	IS	Cl - DS	-0.20	0.66	0.07	-0.35	-0.06	-2.8	83	0.006**	0.853	0.000	-0.30
		Pair 2		D - MSA	0.35	0.96	0.11	0.14	0.57	3.329	81	0.001***	0.753	0.000	0.36
		Pair 3		MSA - Cl	-0.14	1.00	0.11	-0.36	0.09	-1.22	80	0.226	0.72	0.000	
	1	Pair 1	IS	Cl - DS	-0.12	0.86	0.06	-0.25	0.00	-1.94	178	0.054	0.778	0.000	
		Pair 2		D - MSA	0.48	1.04	0.08	0.32	0.63	6.149	178	0.000***	0.74	0.000	0.46
		Pair 3		MSA - Cl	-0.35	1.05	0.08	-0.51	-0.20	-4.49	178	0.000***	0.72	0.000	-0.33
	0	Pair 1	OTS	Cl - DS	0.24	0.76	0.08	0.08	0.41	2.941	83	0.004**	0.79	0.000	0.32
		Pair 2		D - MSA	0.26	0.81	0.09	0.08	0.44	2.932	80	0.004**	0.774	0.000	0.32
		Pair 3		MSA - Cl	-0.46	0.90	0.10	-0.66	-0.26	-4.65	80	0.000***	0.716	0.000	-0.51
	1	Pair 1	OTS	Cl - DS	0.18	0.75	0.06	0.06	0.29	3.126	178	0.002**	0.79	0.000	0.24
		Pair 2		D - MSA	0.38	0.77	0.06	0.27	0.50	6.609	178	0.000***	0.786	0.000	0.49
		Pair 3		MSA - Cl	-0.56	0.90	0.07	-0.69	-0.42	-8.27	178	0.000***	0.686	0.000	-0.62

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

Paired Samples Test – Gender Split File (0 = Male; 1 = Female)																
Gender				Mean	SD	SE	95% CI	t	df	Sig. (2-tailed)	Corr.	Corr Sig.	Effect Size			
T2	0	Pair 1	CS	Cl - DS	-0.22	0.99	0.13	-0.48	0.04	-1.67	55	0.101	0.734	0.000	0.58	
		Pair 2		D - MSA	0.55	0.95	0.13	0.29	0.80	4.267	54	0.000***	0.783	0.000		
		Pair 3		MSA - Cl	-0.32	1.01	0.14	-0.60	-0.05	-2.34	53	0.023	0.749	0.000		
	1	Pair 1		Cl - DS	0.05	0.95	0.09	-0.13	0.22	0.524	114	0.602	0.751	0.000		
		Pair 2		D - MSA	0.13	0.88	0.08	-0.03	0.30	1.576	111	0.118	0.792	0.000		
		Pair 3		MSA - Cl	-0.19	0.93	0.09	-0.36	-0.01	-2.14	111	0.034	0.771	0.000		
	0	Pair 1		IS	Cl - DS	-0.07	1.01	0.13	-0.34	0.20	-0.53	56	0.602	0.758	0.000	0.39
		Pair 2			D - MSA	0.47	1.22	0.16	0.14	0.80	2.833	54	0.006**	0.702	0.000	
		Pair 3			MSA - Cl	-0.41	1.42	0.19	-0.79	-0.02	-2.12	54	0.038	0.597	0.000	
	1	Pair 1	Cl - DS		-0.08	0.97	0.09	-0.26	0.10	-0.93	114	0.355	0.797	0.000	0.35	
		Pair 2	D - MSA		0.41	1.17	0.11	0.19	0.63	3.688	111	0.000***	0.723	0.000		
		Pair 3	MSA - Cl		-0.35	0.94	0.09	-0.52	-0.17	-3.91	111	0.000***	0.82	0.000		-0.37
	0	Pair 1	OTS		Cl - DS	-0.01	0.99	0.13	-0.27	0.26	-0.07	55	0.947	0.695	0.000	0.62
		Pair 2			D - MSA	0.44	0.71	0.10	0.24	0.63	4.499	53	0.000***	0.855	0.000	
		Pair 3			MSA - Cl	-0.44	1.14	0.16	-0.76	-0.13	-2.87	53	0.006**	0.617	0.000	
	1	Pair 1		Cl - DS	0.09	0.70	0.07	-0.04	0.22	1.363	114	0.175	0.826	0.000	0.33	
		Pair 2		D - MSA	0.34	1.04	0.10	0.14	0.53	3.459	111	0.001	0.658	0.000		
		Pair 3		MSA - Cl	-0.45	1.01	0.10	-0.64	-0.26	-4.7	111	0.000***	0.648	0.000		-0.45

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

C.20. Self-Concept Descriptive Statistics – Heritage Background Differences

	Variable	N	Min.	Max.	Mean	Std. Error	Std. Deviation	Skewness	Std. Error	z-skew	Kurtosis	Std. Error	z-kurt
NHL	CI CS	201	1.00	6.00	3.464	0.089	1.260	-0.03	0.172	-0.169	-0.694	0.341	-2.035
	CI IS	201	1.00	6.00	4.159	0.089	1.261	-0.44	0.172	-2.570	-0.580	0.341	-1.701
	CI OTS	201	1.00	6.00	3.087	0.085	1.199	0.324	0.172	1.884	-0.442	0.341	-1.296
	D CS	201	1.00	6.00	3.269	0.090	1.275	0.167	0.172	0.971	-0.685	0.341	-2.009
	D IS	201	1.00	6.00	4.338	0.095	1.350	-0.54	0.172	-3.140	-0.493	0.341	-1.446
	D OTS	201	1.00	6.00	2.876	0.088	1.252	0.559	0.172	3.250	-0.242	0.341	-0.710
	MSA CS	199	1.00	6.00	2.935	0.095	1.334	0.36	0.172	2.093	-0.486	0.343	-1.417
	MSA IS	199	1.00	6.00	3.879	0.109	1.531	-0.3	0.172	-1.756	-0.999	0.343	-2.913
	MSA OTS	199	1.00	6.00	2.570	0.087	1.231	0.557	0.172	3.238	-0.414	0.343	-1.207
	CI CS T2	128	1.00	5.67	2.945	0.103	1.161	0.009	0.214	0.042	-0.852	0.425	-2.005
	CI IS T2	128	1.00	6.00	3.737	0.134	1.514	-0.35	0.214	-1.621	-0.911	0.425	-2.144
	CI OTS T2	128	0.75	5.75	2.789	0.101	1.137	0.285	0.214	1.332	-0.424	0.425	-0.998
	D CS T2	127	1.00	5.67	2.969	0.108	1.217	-0.01	0.215	-0.023	-0.753	0.427	-1.763
	D IS T2	127	1.00	6.00	3.850	0.140	1.581	-0.36	0.215	-1.674	-0.870	0.427	-2.037
	D OTS T2	127	1.00	6.00	2.685	0.108	1.213	0.59	0.215	2.744	0.027	0.427	0.063
	MSA CS T2	124	1.00	5.67	2.610	0.107	1.188	0.15	0.217	0.691	-0.974	0.431	-2.260
	MSA IS T2	124	1.00	6.00	3.352	0.141	1.568	-0.02	0.217	-0.097	-1.077	0.431	-2.499
	MSA OTS T2	124	0.75	6.00	2.343	0.109	1.209	0.824	0.217	3.797	0.184	0.431	0.427

	Variable	N	Min.	Max.	Mean	Std. Error	Std. Deviation	Skewness	Std. Error	z-skew	Kurtosis	Std. Error	z-kurt
HL	Cl CS	64	0.33	6.00	3.672	0.156	1.247	-0.4	0.299	-1.338	-0.046	0.590	-0.078
	Cl IS	62	2.00	6.00	4.296	0.132	1.037	-0.32	0.304	-1.066	-0.854	0.599	-1.426
	Cl OTS	63	1.25	5.25	2.988	0.118	0.939	0.167	0.302	0.553	-0.465	0.595	-0.782
	D CS	62	1.00	6.00	3.581	0.171	1.343	0.035	0.304	0.115	-0.771	0.599	-1.287
	D IS	64	1.67	6.00	4.276	0.157	1.255	-0.39	0.299	-1.318	-0.894	0.590	-1.515
	D OTS	62	1.00	5.25	2.863	0.122	0.962	0.299	0.304	0.984	-0.205	0.599	-0.342
	MSA CS	62	1.00	6.00	3.194	0.181	1.423	0.231	0.304	0.760	-0.790	0.599	-1.319
	MSA IS	62	1.00	6.00	3.935	0.173	1.363	-0.45	0.304	-1.464	-0.638	0.599	-1.065
	MSA OTS	61	1.00	5.25	2.439	0.133	1.039	0.425	0.306	1.389	-0.313	0.604	-0.518
	Cl CS T2	46	1.00	6.00	4.130	0.210	1.427	-0.63	0.35	-1.800	-0.598	0.688	-0.869
	Cl IS T2	46	1.00	6.00	4.500	0.187	1.270	-0.85	0.35	-2.414	0.488	0.688	0.709
	Cl OTS T2	47	1.00	6.00	3.133	0.180	1.235	0.29	0.347	0.836	-0.479	0.681	-0.703
	D CS T2	45	1.00	6.00	4.215	0.208	1.395	-0.5	0.354	-1.415	-0.412	0.695	-0.593
	D IS T2	45	1.33	6.00	4.496	0.188	1.263	-0.43	0.354	-1.220	-0.835	0.695	-1.201
	D OTS T2	45	1.25	6.00	3.261	0.194	1.300	0.779	0.354	2.201	-0.230	0.695	-0.331
	MSA CS T2	43	1.00	6.00	4.109	0.220	1.440	-0.29	0.361	-0.806	-0.825	0.709	-1.164
	MSA IS T2	43	1.00	6.00	4.248	0.249	1.632	-0.79	0.361	-2.183	-0.511	0.709	-0.721
	MSA OTS T2	43	1.00	6.00	2.826	0.228	1.494	0.661	0.361	1.831	-0.315	0.709	-0.444

C.21. T-Test Results – Between Group Comparisons of Self-Concept scores – Heritage Background

		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		Effect Size
									Lower	Upper	
T1	CI CS	1.133	0.288	-1.397	257	0.164	-0.253	0.181	-0.609	0.104	
	CI IS	2.662	0.104	-0.871	119.61	0.386	-0.140	0.161	-0.459	0.179	
	CI OTS	5.158	0.024	0.697	257	0.486	0.116	0.167	-0.212	0.445	
	D CS	0.046	0.829	-1.619	96.034	0.109	-0.314	0.194	-0.699	0.071	
	D IS	0.184	0.668	0.366	110.05	0.715	0.068	0.185	-0.300	0.435	
	D OTS	7.8	0.006	0.281	256	0.779	0.049	0.175	-0.295	0.393	
	MSA CS	0.578	0.448	-1.25	95.254	0.215	-0.258	0.206	-0.667	0.152	
	MSA IS	2.11	0.148	-0.152	110.71	0.879	-0.032	0.207	-0.442	0.379	
	MSA OTS	2.491	0.116	0.942	113.69	0.348	0.152	0.162	-0.168	0.472	
T2	CI CS T2	2.558	0.112	-5.348	67.174	0.000***	-1.253	0.234	-1.721	-0.785	-.98
	CI IS T2	3.963	0.048	-3.181	168	0.002**	-0.799	0.251	-1.295	-0.303	-.58
	CI OTS T2	0.501	0.48	-1.742	75.224	0.086	-0.363	0.208	-0.777	0.052	
	D CS T2	1.385	0.241	-5.3	66.82	0.000***	-1.266	0.239	-1.743	-0.789	-.97
	D IS T2	1.816	0.18	-2.632	93.921	0.010	-0.626	0.238	-1.098	-0.154	
	D OTS T2	0.105	0.746	-2.752	71.97	0.007	-0.616	0.224	-1.062	-0.170	
	MSA CS T2	4.273	0.04	-6.887	162	0.000***	-1.539	0.223	-1.980	-1.098	-1.18
	MSA IS T2	0.01	0.921	-3.409	71.078	0.001***	-0.957	0.281	-1.516	-0.397	-.61
	MSA OTS T2	2.09	0.15	-2.016	60.812	0.048	-0.512	0.254	-1.019	-0.004	

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

C.22. T-Test Results – Heritage Split Comparison of Motivation Scale Average-Score Means

Paired Samples Test – Heritage Background Split File (0 = NHL; 1 = HL)															
					Mean	SD	SE	95% CI		t	df	Sig. (2-tailed)	Corr.	Corr Sig.	Effect Size
								Lower	Upper						
T1	0	Pair 1	CS	Cl - DS	0.20	0.88	0.06	0.07	0.32	3.165	200	0.002**	0.761	0.000	.23
		Pair 2		D - MSA	0.34	0.81	0.06	0.22	0.45	5.877	198	0.000***	0.809	0.000	.42
		Pair 3		MSA - Cl	-0.52	0.89	0.06	-0.65	-0.40	-8.23	198	0.000***	0.763	0.000	-.58
	1	Pair 1	CS	Cl - DS	0.15	0.92	0.12	-0.09	0.38	1.239	61	0.220	0.741	0.000	
		Pair 2		D - MSA	0.33	0.96	0.12	0.08	0.57	2.669	60	0.010*	0.757	0.000	.34
		Pair 3		MSA - Cl	-0.52	1.06	0.13	-0.79	-0.25	-3.826	61	0.000***	0.682	0.000	-.49
	0	Pair 1	IS	Cl - DS	-0.18	0.80	0.06	-0.29	-0.07	-3.18	200	0.002**	0.815	0.000	-.23
		Pair 2		D - MSA	0.47	1.01	0.07	0.32	0.61	6.472	198	0.000***	0.759	0.000	.47
		Pair 3		MSA - Cl	-0.28	1.08	0.08	-0.43	-0.13	-3.645	198	0.000***	0.719	0.000	-.26
	1	Pair 1	IS	Cl - DS	-0.05	0.81	0.10	-0.26	0.15	-0.522	61	0.603	0.749	0.000	
		Pair 2		D - MSA	0.35	1.01	0.13	0.09	0.61	2.713	61	0.009	0.697	0.000	
		Pair 3		MSA - Cl	-0.31	0.91	0.12	-0.54	-0.07	-2.622	60	0.011	0.735	0.000	
	0	Pair 1	OTS	Cl - DS	0.21	0.78	0.05	0.10	0.32	3.856	200	0.000***	0.8	0.000	.27
		Pair 2		D - MSA	0.32	0.76	0.05	0.21	0.42	5.878	198	0.000***	0.815	0.000	.42
		Pair 3		MSA - Cl	-0.51	0.89	0.06	-0.63	-0.39	-8.084	198	0.000***	0.729	0.000	-.57
	1	Pair 1	OTS	Cl - DS	0.15	0.68	0.09	-0.02	0.33	1.775	61	0.081	0.74	0.000	
		Pair 2		D - MSA	0.44	0.86	0.11	0.22	0.66	3.973	60	0.000***	0.632	0.000	.51
		Pair 3		MSA - Cl	-0.59	0.93	0.12	-0.83	-0.35	-4.906	60	0.000***	0.554	0.000	-.63

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

Paired Samples Test – Heritage Background Split File (0 = NHL; 1 = HL)															
					Mean	SD	SE	95% CI		t	df	Sig. (2-tailed)	Corr.	Corr Sig.	Effect Size
								Lower	Upper						
T2	0	Pair 1	CS	CI - DS	-0.02	0.93	0.08	-0.18	0.15	-0.224	125	0.823	0.694	0.000	
		Pair 2		D - MSA	0.33	0.79	0.07	0.19	0.47	4.702	123	0.000***	0.784	0.000	.42
		Pair 3		MSA - CI	-0.32	0.95	0.09	-0.49	-0.15	-3.778	122	0.000***	0.679	0.000	-.34
		Pair 1		CI - DS	-0.10	1.08	0.16	-0.43	0.22	-0.646	44	0.522	0.711	0.000	
		Pair 2		D - MSA	0.08	1.21	0.19	-0.30	0.45	0.419	42	0.678	0.638	0.000	
		Pair 3		MSA - CI	0.03	0.93	0.14	-0.26	0.32	0.218	42	0.828	0.794	0.000	
	1	Pair 1	IS	CI - DS	-0.10	0.92	0.08	-0.26	0.06	-1.26	126	0.210	0.826	0.000	
		Pair 2		D - MSA	0.49	1.10	0.10	0.30	0.69	4.99	123	0.000***	0.751	0.000	.45
		Pair 3		MSA - CI	-0.40	1.12	0.10	-0.60	-0.20	-3.969	123	0.000***	0.734	0.000	-.36
		Pair 1		CI - DS	-0.01	1.15	0.17	-0.36	0.33	-0.086	44	0.932	0.587	0.000	
		Pair 2		D - MSA	0.23	1.38	0.21	-0.19	0.66	1.101	42	0.277	0.57	0.000	
		Pair 3		MSA - CI	-0.26	1.09	0.17	-0.60	0.07	-1.582	42	0.121	0.745	0.000	
	0	Pair 1	OTS	CI - DS	0.12	0.72	0.06	-0.01	0.24	1.802	125	0.074	0.814	0.000	
		Pair 2		D - MSA	0.36	0.86	0.08	0.20	0.51	4.612	122	0.000***	0.751	0.000	.42
		Pair 3		MSA - CI	-0.47	0.96	0.09	-0.64	-0.30	-5.436	122	0.000***	0.667	0.000	-.49
		Pair 1		CI - DS	-0.11	1.01	0.15	-0.41	0.20	-0.702	44	0.487	0.679	0.000	
		Pair 2		D - MSA	0.41	1.17	0.18	0.05	0.77	2.316	42	0.025	0.659	0.000	
		Pair 3		MSA - CI	-0.39	1.30	0.20	-0.79	0.01	-1.966	42	0.056	0.553	0.000	

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

Appendix C.23 – Interview References to Different Types of Arabic

MSA

1. the stuff we're learning right now, it's more for like *everyday kind of cultural stuff*
2. *normal Arabic*
3. I do think that they're completely different things. That the-- both Arabics are completely different. *Because Arabic has changed.*
4. Researcher: Do you know what kind of Arabic they speak to you in?
Student 1: Um, no, but it's like the same thing as, like, we do in school. I don't know if there's a different dialect...
Student 2: *Fuṣḥā?*
Researcher: They speak to you in fuṣḥā?
Student 2: ... that's the one [we're learning]
5. *school Arabic*
6. The Qur'an is different to the Arabic that you do learn in school, which is more... Um... *For worldly use.*
7. Arabic in school is, like, *conversation type.*
8. This is more, like, *general Arabic*
9. the Arabic *is different to mosque Arabic*-- like, what we learn in mosque.

Classical

1. the Qur'an has like *a different sort of Arabic*
2. *familiar language to, like, Muslims*
3. *Arabic with Islam*
4. the type of Arabic we learn in mosque is *all to do with the Qur'an*
5. *Islamic related*
6. *the classical Arabic*
7. *the Arabic in the Qur'an was completely different from the Arabic language we're studying, and this was much, much harder as well*
8. But you know, *the things that we do in mosque, they don't exactly come in the Arabic here.* Like, it's more like, different
9. the Qur'an in itself is-- the Arabic itself is a bit *different* and it's, like, *a bit more challenging to understand*
10. The Qur'an, because it's *written a bit different*
11. the Arabic in the Qur'an is *completely different to the Arabic that people speak, like you, you wouldn't speak it*
12. 'Alimah Arabic is more, like, *religion-based.*
13. If I go that path, I need, like, the-- the, the one, *the Arabic that's in the Qur'an and scriptures and stuff like that*-- That's more of a[n] *ancient Arabic.* And I need to understand and that means that I have to learn Arabic all over again, cos it's not the same
14. She says that Arabic-- the Arabic is different to *mosque Arabic*-- like, what we learn in mosque. Two different types of Arabic.

Dialect

1. *Algerian dialect*
2. I think I'm perfectly fine with *my Arabic dialects*.
3. But as Arabic is just-- just different ways of Arabic, like, *colloquial*, and different stuff, so I do-- I do wish that we learnt more different styles of Arabic. Not just one way. But like, um, like, even, like, in Arab countries, a lot of people talk in *slang Arabic* as well, different ways, so it'd be helpful if we learnt all types.
4. I just want to get back and I want -- in a few years, I would like to know the difference when I'm talking *dialect* and normal-- Um, *fuṣḥā* Arabic. Because, I don't know, I can't really-- I mix both of them together and I get really confused.
5. Student 1: Uh, I think *dialect Arabic* I'll probably use more with my family, because none of them speak, like, *fuṣḥā*.
Researcher: Okay.
Student 1: And... Not only with my family, but, like, if I go back to-- back home... Or with other people who speak dialect Arabic only... And I think *fuṣḥā*, I'll probably use it... With non-dialect speakers? Like, if they don't speak dialect, then I'll probably speak *fuṣḥā*.
6. I think I need the *dialect version* if I go to the other path.
7. I think that helps me a lot, because I-- even though I still don't know how to differentiate properly, it does help me kind of grasp *dialects* and *fuṣḥā* Arabic. Because I-- I know, like, some of the words can be very similar. But also very different. So, words that are similar, I picked up on, and then words that were different, I picked up on as well. Which helps me differentiate between both of the *different Arabics*.

C.24. T-Test Results – Longitudinal Changes in Self-Concept scores – Gender

Average-Score Variables			Mean	SD	SE	95% Confidence Interval		t	df	Sig. (2-tailed)
						Lower	Upper			
M	Pair 1	D CS - D CS T2	0.358	1.644	0.222	-0.087	0.802	1.613	54	0.113
	Pair 2	D IS - D IS T2	0.133	1.850	0.249	-0.367	0.633	0.535	54	0.595
	Pair 3	D OTS - D OTS T2	0.210	1.740	0.232	-0.256	0.676	0.903	55	0.371
	Pair 1	Cl CS - Cl CS T2	0.068	1.823	0.248	-0.430	0.565	0.274	53	0.785
	Pair 2	Cl IS - Cl IS T2	0.113	1.934	0.258	-0.405	0.631	0.438	55	0.663
	Pair 3	Cl OTS - Cl OTS T2	0.028	1.802	0.245	-0.464	0.520	0.113	53	0.910
	Pair 1	MSA CS - MSA CS T2	0.045	1.944	0.270	-0.496	0.586	0.166	51	0.868
	Pair 2	MSA IS - MSA IS T2	0.295	2.127	0.295	-0.297	0.887	1.000	51	0.322
	Pair 3	MSA OTS - MSA OTS T2	0.025	1.514	0.214	-0.405	0.455	0.117	49	0.908
F	Pair 1	D CS - D CS T2	0.316	1.525	0.142	0.036	0.597	2.232	115	0.028
	Pair 2	D IS - D IS T2	0.345	1.696	0.157	0.033	0.657	2.19	115	0.031
	Pair 3	D OTS - D OTS T2	0.205	1.541	0.143	-0.079	0.488	1.431	115	0.155
	Pair 1	Cl CS - Cl CS T2	0.212	1.590	0.148	-0.082	0.505	1.427	114	0.156
	Pair 2	Cl IS - Cl IS T2	0.412	1.897	0.177	0.061	0.762	2.327	114	0.022
	Pair 3	Cl OTS - Cl OTS T2	0.141	1.691	0.158	-0.171	0.454	0.896	114	0.372
	Pair 1	MSA CS - MSA CS T2	-0.027	1.647	0.156	-0.335	0.282	-0.172	111	0.864
	Pair 2	MSA IS - MSA IS T2	0.277	2.021	0.191	-0.102	0.655	1.449	111	0.150
	Pair 3	MSA OTS - MSA OTS T2	0.114	1.655	0.156	-0.196	0.424	0.728	111	0.468

C.25. T-Test Results – Longitudinal Changes in Self-Concept Scores – Heritage Backgrounds

Average-Score Variables			Mean	SD	SE	95% Confidence Interval		t	df	Sig. (2-tailed)	Effect Size
						Lower	Upper				
NHL	Pair 1	D CS - D CS T2	0.378	1.600	0.143	0.096	0.660	2.655	125	0.009**	.24
	Pair 2	D IS - D IS T2	0.492	1.997	0.178	0.140	0.844	2.766	125	0.007**	.25
	Pair 3	D OTS - D OTS T2	0.198	1.697	0.151	-0.101	0.498	1.313	125	0.192	
	Pair 1	Cl CS - Cl CS T2	0.540	1.510	0.134	0.274	0.806	4.013	125	0.000***	.36
	Pair 2	Cl IS - Cl IS T2	0.409	1.827	0.162	0.089	0.730	2.526	126	0.013*	.23
	Pair 3	Cl OTS - Cl OTS T2	0.284	1.613	0.144	-0.001	0.568	1.974	125	0.051	
	Pair 1	MSA CS - MSA CS T2	0.301	1.672	0.151	0.002	0.599	1.995	122	0.048	
	Pair 2	MSA IS - MSA IS T2	0.415	2.174	0.196	0.027	0.803	2.115	122	0.036	
	Pair 3	MSA OTS - MSA OTS T2	0.180	1.615	0.146	-0.109	0.470	1.233	121	0.220	
HL	Pair 1	D CS - D CS T2	-0.457	1.711	0.261	-0.984	0.069	-1.753	42	0.087	
	Pair 2	D IS - D IS T2	-0.185	1.548	0.231	-0.650	0.280	-0.802	44	0.427	
	Pair 3	D OTS - D OTS T2	-0.169	1.788	0.273	-0.719	0.382	-0.618	42	0.540	
	Pair 1	Cl CS - Cl CS T2	-0.259	1.563	0.233	-0.729	0.210	-1.113	44	0.272	
	Pair 2	Cl IS - Cl IS T2	-0.106	1.430	0.216	-0.541	0.329	-0.492	43	0.625	
	Pair 3	Cl OTS - Cl OTS T2	-0.005	1.572	0.232	-0.472	0.461	-0.023	45	0.981	
	Pair 1	MSA CS - MSA CS T2	-0.919	1.636	0.256	-1.435	-0.402	-3.596	40	0.001***	-.56
	Pair 2	MSA IS - MSA IS T2	-0.114	1.575	0.246	-0.611	0.383	-0.463	40	0.646	
	Pair 3	MSA OTS - MSA OTS T2	-0.200	1.572	0.249	-0.703	0.303	-0.805	39	0.426	

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

C.26. T-Test Results – Proficiency Tests Changes Over Time

Paired Samples Test		Mean	SD	SE	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Y/N T1 - Y/N T2	0.12426	13.948	1.0729	-1.99384	2.24236	0.116	168	0.908
Pair 2	Cloze T1 - Cloze T2	-0.5697	4.987	0.38823	-1.33626	0.19687	-1.467	164	0.144

C.27. Mean Comparison – Self-Rated Proficiency Over Time

Related-Samples Wilcoxon Signed Rank Test	MSA Writing	MSA Speaking	MSA Reading Comprehension	MSA Listening Comprehension	Dialect Speaking	Dialect Listening Comprehension	MSA v. Dialect	Quran Listening Comprehension	Quran Reading Comprehension
Total N	171	170	170	169	169	168	169	169	169
Test Statistic	2857.5	4516.5	4476.5	3519	4037	3640	4656.5	4546.5	4097.5
Standard Error	319.64	417.681	422.386	375.799	446.422	436.784	471.829	440.138	421.706
Standardized Test Statistic	-0.267	0.776	0.518	-0.295	-1.088	-1.715	-0.442	0.207	-0.225
Asymptotic Sig.(2-sided test)	0.789	0.438	0.604	0.768	0.277	0.086	0.659	0.836	0.822

C.28. Proficiency Test Descriptives – Gender

		N	Min.	Max.	Mean	SE	SD	Skewness	Std. Error	Kurtosis	Std. Error
F	YN Total T1	89	0	60	24.483	1.68	16	0.664	0.255	-0.539	0.506
	Cloze Total T1	85	0	25	13.44	0.65	6	0.101	0.261	-0.692	0.517
	YN Total T2	60	0	54	25.45	1.81	14	0.19	0.309	-0.683	0.608
	Cloze Total T2	59	0	25	15.2	0.83	6.4	-0.165	0.311	-1.051	0.613
M	YN Total T1	173	0	60	24.012	1.17	15	0.687	0.185	-0.404	0.367
	Cloze Total T1	178	0	25	13.97	0.47	6.3	0.176	0.182	-0.94	0.362
	YN Total T2	111	0	59	22.61	1.23	13	0.33	0.229	-0.25	0.455
	Cloze Total T2	111	0	25	15.6	0.66	7	-0.531	0.229	-0.549	0.455

C.29. T-Test Results – Proficiency Test Score Comparisons between Genders

	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	SE Difference	Lower	Upper
Y/N T1	0.407	0.524	0.231	173.23	0.818	0.472	2.044	-3.563	4.507
Y/N T2	0.357	0.551	1.323	114.45	0.189	3.240	2.449	-1.612	8.092
Cloze T1	0.313	0.576	-0.668	178.22	0.505	-0.546	0.817	-2.159	1.067
Cloze T2	0.078	0.78	-0.376	127.61	0.708	-0.400	1.065	-2.508	1.708

C.30. T-Test Results – Proficiency Test Score between time comparisons within Genders

Paired Samples Test			Mean	SD	SE	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
						Lower	Upper			
M	Pair 1	Y/N T1 - Y/N T2	-2.667	12.898	1.665	-5.999	0.665	-1.601	59	0.115
	Pair 2	Cloze T1 - Cloze T2	-0.864	4.058	0.528	-1.922	0.193	-1.636	58	0.107
F	Pair 1	Y/N T1 - Y/N T2	1.661	14.319	1.371	-1.058	4.379	1.211	108	0.229
	Pair 2	Cloze T1 - Cloze T2	-0.406	5.447	0.529	-1.455	0.643	-0.767	105	0.445

C.31. Self-Rated Proficiency Descriptive Statistics - Gender

		Males								Females							
		N	Mean	SE	SD	Skew	SE	Kurt	SE	N	Mean	SE	SD	Skew	SE	Kurt	SE
T1	Reading Words	86	5.4	0.11	1	-2.04	0.26	4.77	0.51	179	5.32	0.07	0.92	-1.45	0.18	2.32	0.36
	MSA (Writing)	85	5.15	0.13	1.23	-1.64	0.26	2.39	0.52	179	5.04	0.09	1.15	-1.23	0.18	1.12	0.36
	MSA (Speaking)	85	4.41	0.15	1.39	-0.78	0.26	0.02	0.52	179	4.5	0.1	1.27	-0.51	0.18	-0.41	0.36
	MSA (Reading Comprehension)	85	4.61	0.15	1.38	-0.85	0.26	0.07	0.52	179	4.54	0.1	1.28	-0.78	0.18	0.04	0.36
	MSA (Listening Comprehension)	86	4.66	0.15	1.4	-1	0.26	0.32	0.51	179	4.6	0.09	1.21	-0.76	0.18	0.52	0.36
	MSA v. Dialect	85	3.69	0.19	1.79	-0.05	0.26	-1.4	0.52	179	3.24	0.14	1.8	0.21	0.18	-1.27	0.36
	Dialect (Speaking)	84	3.48	0.19	1.78	0.06	0.26	-1.3	0.52	179	2.97	0.12	1.62	0.28	0.18	-0.97	0.36
	Dialect (Listening Comprehension)	84	3.92	0.18	1.64	-0.16	0.26	-1.15	0.52	179	3.47	0.13	1.67	-0.07	0.18	-1.17	0.36
	Quran (Listening Comprehension)	85	3.92	0.17	1.56	-0.21	0.26	-1.02	0.52	179	3.63	0.12	1.65	-0.03	0.18	-1.13	0.36
	(Word Reading with Diacritics)	86	5.78	0.08	0.74	-3.51	0.26	12	0.51	179	5.58	0.08	1.05	-2.77	0.18	7.34	0.36
	Quran (Reading Comprehension)	86	3.71	0.17	1.59	-0.15	0.26	-1.07	0.51	179	3.54	0.12	1.62	-0.04	0.18	-1.1	0.36

		Males								Females							
		N	Mean	SE	SD	Skew	SE	Kurt	SE	N	Mean	SE	SD	Skew	SE	Kurt	SE
T2	Reading Words	57	5.37	0.14	1.08	-2.21	0.32	5.26	0.62	115	5.13	0.11	1.19	-1.66	0.23	2.47	0.45
	MSA (Writing)	57	5.32	0.14	1.04	-1.87	0.32	4.3	0.62	115	4.87	0.12	1.32	-1.38	0.23	1.55	0.45
	MSA (Speaking)	57	5.05	0.17	1.3	-1.46	0.32	1.48	0.62	114	4.39	0.14	1.47	-0.73	0.23	-0.38	0.45
	MSA (Reading Comprehension)	57	4.84	0.17	1.28	-0.92	0.32	0.06	0.62	114	4.49	0.13	1.35	-0.76	0.23	-0.19	0.45
	MSA (Listening Comprehension)	56	4.82	0.16	1.19	-0.91	0.32	0.63	0.63	114	4.48	0.13	1.38	-0.9	0.23	0.22	0.45
	MSA v. Dialect	56	3.7	0.25	1.85	-0.18	0.32	-1.35	0.63	114	3.25	0.17	1.78	0.21	0.23	-1.28	0.45
	Dialect (Speaking)	56	3.07	0.25	1.89	0.3	0.32	-1.34	0.63	114	2.86	0.16	1.72	0.44	0.23	-1.08	0.45
	Dialect (Listening Comprehension)	56	3.5	0.26	1.92	-0.08	0.32	-1.45	0.63	114	3.25	0.15	1.61	0.08	0.23	-1.17	0.45
	Quran (Listening Comprehension)	56	3.68	0.21	1.57	-0.2	0.32	-1.01	0.63	114	3.75	0.16	1.7	-0.18	0.23	-1.13	0.45
	(Word Reading with Diacritics)	56	5.64	0.12	0.9	-3.21	0.32	12.4	0.63	114	5.48	0.1	1.08	-2.33	0.23	5.32	0.45
	Quran (Reading Comprehension)	56	3.55	0.22	1.67	0.02	0.32	-1.21	0.63	114	3.49	0.16	1.72	0.07	0.23	-1.17	0.45

C.32. Mean Comparison Results – Self-Rated Proficiency Comparisons by Gender

Independent-Samples Mann-Whitney U Test		MSA Writing	MSA Speaking	MSA Reading Comprehension	MSA Listening Comprehension	Dialect Speaking	Dialect Listening Comprehension	MSA v. Dialect	Quran Listening Comprehension	Quran Reading Comprehension
T1	Total N	264	264	264	265	263	263	264	264	265
	Mann-Whitney U	6952.5	7739	7231.5	7171.5	6319.5	6416.5	6531	6843	7241
	Wilcoxon W	23062.5	23849	23341.5	23281.5	22429.5	22526.5	22641	22953	23351
	Test Statistic	6952.5	7739	7231.5	7171.5	6319.5	6416.5	6531	6843	7241
	Standard Error	537.262	562.847	561.216	564.314	564.019	566.263	570.263	570.593	575.422
	Standardized Test Statistic	-1.219	0.234	-0.67	-0.931	-2.125	-1.945	-1.888	-1.34	-0.792
	Asymptotic Sig.(2-sided test)	0.223	0.815	0.503	0.352	0.034	0.052	0.059	0.18	0.428
T2	Total N	172	171	171	170	170	170	170	170	170
	Mann-Whitney U	2597.5	2333	2730.5	2778	3012.5	2947	2743.5	3268	3118.5
	Wilcoxon W	9267.5	8888	9285.5	9333	9567.5	9502	9298.5	9823	9673.5
	Test Statistic	2597.5	2333	2730.5	2778	3012.5	2947	2743.5	3268	3118.5
	Standard Error	287.397	294.203	295.138	291.667	293.788	296.946	296.848	296.787	296.498
	Standardized Test Statistic	-2.366	-3.113	-1.757	-1.419	-0.611	-0.825	-1.511	0.256	-0.248
	Asymptotic Sig.(2-sided test)	0.018	0.002*	0.079	0.156	0.541	0.409	0.131	0.798	0.804

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

C.33. Mean Comparison Results – Self-Rated Proficiency between time comparisons within Gender

Related-Samples Wilcoxon Signed Rank Test		MSA Writing	MSA Speaking	MSA Reading Comprehension	MSA Listening Comprehension	Dialect Speaking	Dialect Listening Comprehension	MSA v. Dialect	Quran Listening Comprehension	Quran Reading Comprehension
M	Total N	56	56	56	55	55	54	55	55	55
	Test Statistic	297	554	419.5	509	433	429.5	523	449	469
	Standard Error	52.554	70.707	70.708	81.258	90.903	93.861	93.802	86.801	82.041
	Standardized Test Statistic	0.628	2.319	0.417	0.443	-1.183	-1.433	-0.437	-0.789	-0.049
	Asymptotic Sig.(2-sided test)	0.53	0.02	0.677	0.658	0.237	0.152	0.662	0.43	0.961
F	Total N	115	114	114	114	114	114	114	114	114
	Test Statistic	1330	1873	2179.5	1372	1844	1584.5	2065	2124.5	1815.5
	Standard Error	188.96	243.28	247.27	194.19	237.94	226.1	254.59	238.15	230.17
	Standardized Test Statistic	-0.704	-0.717	0.35	-0.667	-0.479	-1.075	-0.291	0.699	-0.239
	Asymptotic Sig.(2-sided test)	0.482	0.473	0.726	0.505	0.632	0.282	0.771	0.484	0.811

C.34. Proficiency Test Descriptive Statistics – Heritage Background

		N	Min.	Max.	Mean	SE	SD	Skewness	Std. Error	Kurtosis	Std. Error
NHL	YN Total T1	197	0	52	19.234	0.83	12	0.775	0.173	0.197	0.345
	Cloze Total T1	199	0	25	13.29	0.43	6.1	0.22	0.172	-0.73	0.343
	YN Total T2	125	0	48	20.89	1.05	12	0.21	0.217	-0.736	0.43
	Cloze Total T2	124	0	25	14.48	0.58	6.4	-0.29	0.217	-0.686	0.431
HL	YN Total T1	65	0	60	39.139	1.99	16	-0.494	0.297	-0.785	0.586
	Cloze Total T1	64	3	25	15.38	0.78	6.2	-0.037	0.299	-1.148	0.59
	YN Total T2	46	0	59	31	2.2	15	-0.162	0.35	-0.383	0.688
	Cloze Total T2	46	0	25	18.13	1.04	7	-1.054	0.35	0.265	0.688

C.35. T-Test Results – Proficiency Test Score Comparisons between Heritage Backgrounds

	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	SE Difference	Lower	Upper
Y/N T1	18.027	0.000	-10.782	260	0.000***	-19.905	1.846	-23.540	-16.270
Y/N T2	8.015	0.005	-4.825	172	0.000***	-11.655	2.416	-16.423	-6.887
Cloze T1	13.222	0.000	-4.831	262	0.000***	-4.138	0.857	-5.825	-2.451
Cloze T2	0.261	0.610	-3.080	74.579	0.003**	-3.655	1.187	-6.019	-1.291

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

C.36. T-Test Results – Proficiency Test Score between time comparisons within Heritage Backgrounds

Paired Samples Test			Mean	SD	SE	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
						Lower	Upper			
NHL	Pair 1	Y/N T1 - Y/N T2	-1.795	11.694	1.059	-3.891	0.301	-1.695	121	0.093
	Pair 2	Cloze T1 - Cloze T2	-0.367	4.524	0.413	-1.184	0.451	-0.888	119	0.376
HL	Pair 1	Y/N T1 - Y/N T2	5.106	17.766	2.591	-0.110	10.323	1.971	46	0.055
	Pair 2	Cloze T1 - Cloze T2	-1.111	6.076	0.906	-2.937	0.714	-1.227	44	0.226

C.37. Self-Rated Proficiency Descriptive Statistics - Heritage Background

		Heritage								Non-Heritage							
		N	Mean	SE	SD	Skew	SE	Kurt	SE	N	Mean	SE	SD	Skew	SE	Kurt	SE
T1	Reading Words	64	5.56	0.094	0.753	-1.83	0.299	2.983	0.59	201	5.28	0.07	0.986	-1.56	0.172	2.872	0.341
	MSA (Writing)	64	5.27	0.124	0.996	-1.76	0.299	4.25	0.59	200	5.01	0.087	1.226	-1.25	0.172	1.006	0.342
	MSA (Speaking)	64	4.64	0.167	1.338	-0.91	0.299	0.244	0.59	200	4.42	0.091	1.293	-0.54	0.172	-0.27	0.342
	MSA (Reading Comprehension)	63	4.79	0.134	1.065	-1.14	0.302	1.801	0.595	201	4.49	0.097	1.375	-0.69	0.172	-0.28	0.341
	MSA (Listening Comprehension)	64	4.88	0.142	1.134	-0.96	0.299	0.735	0.59	201	4.54	0.092	1.304	-0.8	0.172	0.327	0.341
	MSA v. Dialect	63	3.92	0.229	1.817	-0.28	0.302	-1.32	0.595	201	3.22	0.125	1.778	0.241	0.172	-1.24	0.341
	Dialect (Speaking)	63	3.54	0.203	1.615	-0.01	0.302	-0.93	0.595	200	3.00	0.119	1.688	0.321	0.172	-1.07	0.342
	Dialect (Listening Comprehension)	62	4.00	0.185	1.46	-0.23	0.304	-0.87	0.599	201	3.49	0.121	1.715	-0.02	0.172	-1.22	0.341
	Quran (Listening Comprehension)	63	4.13	0.18	1.431	-0.4	0.302	-0.53	0.595	201	3.59	0.117	1.659	0.019	0.172	-1.18	0.341
	(Word Reading with Diacritics)	64	5.86	0.062	0.5	-4.23	0.299	19.53	0.59	201	5.57	0.075	1.061	-2.68	0.172	6.627	0.341
T2	Quran (Reading Comprehension)	64	4.00	0.19	1.522	-0.36	0.299	-0.76	0.59	201	3.47	0.114	1.616	0.023	0.172	-1.13	0.341
	Reading Words	45	5.44	0.151	1.013	-2.52	0.354	7.85	0.695	127	5.13	0.106	1.195	-1.64	0.215	2.277	0.427
	MSA (Writing)	45	5.47	0.121	0.815	-1.6	0.354	2.152	0.695	127	4.86	0.118	1.332	-1.35	0.215	1.464	0.427
	MSA (Speaking)	45	5.27	0.15	1.009	-1.27	0.354	1.081	0.695	126	4.38	0.135	1.512	-0.73	0.216	-0.5	0.428
	MSA (Reading Comprehension)	45	5.42	0.129	0.866	-1.4	0.354	1.085	0.695	126	4.32	0.12	1.348	-0.6	0.216	-0.41	0.428
	MSA (Listening Comprehension)	44	5.27	0.147	0.973	-1.38	0.357	1.743	0.702	126	4.36	0.121	1.359	-0.79	0.216	0.136	0.428
	MSA v. Dialect	44	4.98	0.214	1.422	-1.38	0.357	1.17	0.702	126	2.84	0.142	1.597	0.42	0.216	-0.95	0.428
	Dialect (Speaking)	44	4.16	0.283	1.879	-0.59	0.357	-1.13	0.702	126	2.50	0.136	1.527	0.63	0.216	-0.69	0.428
	Dialect (Listening Comprehension)	44	4.43	0.253	1.676	-0.79	0.357	-0.56	0.702	126	2.95	0.139	1.564	0.195	0.216	-1.16	0.428
	Quran (Listening Comprehension)	44	4.5	0.204	1.355	-0.41	0.357	-0.71	0.702	126	3.45	0.149	1.671	-0.03	0.216	-1.19	0.428
	(Word Reading with Diacritics)	44	5.64	0.122	0.81	-2.26	0.357	4.211	0.702	126	5.50	0.097	1.086	-2.51	0.216	6.39	0.428
	Quran (Reading Comprehension)	44	4.34	0.218	1.446	-0.25	0.357	-1.01	0.702	126	3.22	0.15	1.687	0.233	0.216	-1.14	0.428

C.38. Mean Comparison Results – Self-Rated Proficiency Comparisons by Heritage Backgrounds

Independent-Samples Mann-Whitney U Test		MSA Writing	MSA Speaking	MSA Reading Comprehension	MSA Listening Comprehension	Dialect Speaking	Dialect Listening Comprehension	MSA v. Dialect	Quran Listening Comprehension	Quran Reading Comprehension
T1	Total N	264	264	264	265	263	263	264	264	265
	Mann-Whitney U	6986.5	7119	7008.5	7354	7426.5	7266.5	7711.5	7540	7645.5
	Wilcoxon W	9066.5	9199	9024.5	9434	9442.5	9219.5	9727.5	9556	9725.5
	Test Statistic	6986.5	7119	7008.5	7354	7426.5	7266.5	7711.5	7540	7645.5
	Standard Error	492.78	516.25	511.99	515.86	516.31	515.52	520.25	520.55	526.02
	Standardized Test Statistic	1.19	1.393	1.322	1.787	2.182	2.009	2.653	2.322	2.307
	Asymptotic Sig.(2-sided test)	0.234	0.164	0.186	0.074	0.029	0.045	0.008	0.02	0.021
T2	Total N	172	171	171	170	170	170	170	170	170
	Mann-Whitney U	3619.5	3823	4247	3916	4142	4117.5	4613	3742.5	3800.5
	Wilcoxon W	4654.5	4858	5282	4906	5132	5107.5	5603	4732.5	4790.5
	Test Statistic	3619.5	3823	4247	3916	4142	4117.5	4613	3742.5	3800.5
	Standard Error	268.35	274.82	275.69	271.8	273.78	276.72	276.63	276.57	276.3
	Standardized Test Statistic	2.84	3.595	5.122	4.209	5.004	4.862	6.655	3.509	3.722
	Asymptotic Sig.(2-sided test)	0.005*	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***

Significance indicators:

* $p \leq .05$ (or Bonferroni-adjusted value)

** $p \leq .01$

*** $p \leq .001$

C.39. Mean Comparison Results – Self-Rated Proficiency between time comparisons within Heritage Backgrounds

Related-Samples Wilcoxon Signed Rank Test		MSA Writing	MSA Speaking	MSA Reading Comprehension	MSA Listening Comprehension	Dialect Speaking	Dialect Listening Comprehension	MSA v. Dialect	Quran Listening Comprehension	Quran Reading Comprehension
NHL	Total N	126	125	125	125	125	125	125	125	125
	Test Statistic	1620	2401	2399	1961	1835	1915.5	2335	2446	2140.5
	Standard Error	216.336	281.576	290.475	260.95	292.304	301.475	323.332	287.915	271.355
	Standardized Test Statistic	-0.569	-0.263	-0.608	-1.04	-2.533	-2.529	-1.88	-0.274	-0.691
	Asymptotic Sig.(2-sided test)	0.57	0.793	0.543	0.298	0.011	0.011	0.06	0.784	0.49
HL	Total N	45	45	45	44	44	43	44	44	44
	Test Statistic	183	335.5	334	231	382	268.5	375.5	338	322.5
	Standard Error	36.107	47.616	44.834	38.555	55.581	45.867	50.614	54.992	55.373
	Standardized Test Statistic	0.568	2.163	2.598	1.439	1.826	1.112	2.519	1.046	0.758
	Asymptotic Sig.(2-sided test)	0.57	0.031	0.009	0.15	0.068	0.266	0.012	0.296	0.448

C.40. Frequencies of References to Node Categories by School

	Nodes	Class 4+5 (n=25)		Class 6 (n=3)		Class 8 (n=12)		Classes 10+11 (n=18)		Class 12 (n=10)		Class 13 (n=6)		Class 16 (n=6)		Class 17 (n=15)		Total (n=95)	
Arabic Future Self	Family	7	28%	2	67%	0	0%	1	6%	0	0%	1	17%	1	17%	0	0%	12	13%
	General Communication	3	12%	2	67%	1	8%	2	11%	2	20%	2	33%	0	0%	1	7%	13	14%
	Job	8	32%	1	33%	0	0%	1	6%	4	40%	3	50%	3	50%	4	27%	24	25%
	Just GCSE	3	12%	1	33%	1	8%	1	6%	0	0%	0	0%	2	33%	0	0%	8	8%
	No Arabic	1	4%	1	33%	11	92%	5	28%	3	30%	0	0%	0	0%	4	27%	25	26%
	Other	0	0%	0	0%	0	0%	0	0%	2	20%	0	0%	0	0%	2	13%	4	4%
	Religion	13	52%	2	67%	2	17%	3	17%	0	0%	1	17%	1	17%	7	47%	29	31%
	Skill	1	4%	0	0%	0	0%	3	17%	5	50%	1	17%	1	17%	4	27%	15	16%
	Travel	13	52%	0	0%	1	8%	7	39%	4	40%	3	50%	5	83%	4	27%	37	39%
Current Self (Feelings)	Apathy	2	8%	1	33%	3	25%	4	22%	0	0%	0	0%	0	0%	0	0%	10	11%
	Despondency	3	12%	0	0%	1	8%	12	67%	6	60%	0	0%	2	33%	8	53%	32	34%
	Dislike	2	8%	1	33%	0	0%	0	0%	0	0%	0	0%	0	0%	2	13%	5	5%
	Enjoyment	4	16%	0	0%	1	8%	1	6%	1	10%	2	33%	4	67%	0	0%	13	14%
	Negative	3	12%	1	33%	5	42%	3	17%	3	30%	1	17%	2	33%	5	33%	23	24%
	Positive	8	32%	3	100%	4	33%	3	17%	1	10%	4	67%	2	33%	5	33%	30	32%
Ought-to Self	Consequences - Future plans affected	5	20%	0	0%	0	0%	1	6%	1	10%	1	17%	1	17%	3	20%	12	13%
	Consequences - Parents	8	32%	3	100%	0	0%	0	0%	1	10%	3	50%	0	0%	7	47%	22	23%
	Consequences - Personal Disappoint.	4	16%	3	100%	1	8%	5	28%	3	30%	2	33%	5	83%	4	27%	27	28%
	General Disappointment	4	16%	0	0%	0	0%	1	6%	0	0%	0	0%	0	0%	0	0%	5	5%
	Compulsory	4	16%	3	100%	3	25%	2	11%	0	0%	0	0%	0	0%	1	7%	13	14%
	Exam Pressure	10	40%	0	0%	2	17%	3	17%	3	30%	2	33%	4	67%	1	7%	25	26%
	Parents Pressure	12	48%	1	33%	0	0%	8	44%	5	50%	3	50%	2	33%	10	67%	41	43%
	Religion Pressure	0	0%	2	67%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	2	2%
	School Pressure	5	20%	1	33%	7	58%	1	6%	6	60%	0	0%	1	17%	8	53%	29	31%
	Sixth Form College	1	4%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	1	7%	2	2%
	OTS - Self	7	28%	2	67%	0	0%	4	22%	4	40%	2	33%	2	33%	4	27%	25	26%

C.41. Notes on Consistency

One area of interest in mixed-methods research is the consistency of findings across the datasets. In order to examine whether or not participants' responses were consistent in the realm of Arabic FS concept, participants who had responded to both surveys and interviews were isolated and their responses examined. In total, 66 of the 95 participants interviewed had also completed both surveys.

Open-ended survey items pertaining to Arabic FS concepts were coded, along with the interview data, into 9 categories, represented in Table 35.1. The respective case frequencies and percentages are listed. What is interesting to note is that while *Travel* clearly emerges as the most frequently cited Arabic FS image in the surveys and interviews, followed by *Religion*, and then *Job*, the number of cases within each is not consistent.

Table 35.1. Comparison of Interview and Survey references to AFS codes						
Future Self-concept	Interview		Survey T1		Survey T2	
	Count	%	Count	%	Count	%
Travel	37	56.1	24	36.4	15	22.7
Family	7	10.6	6	9.1	5	7.6
Religion	22	33.3	11	16.7	11	16.7
Job	16	24.2	11	16.7	10	15.2
Skill	9	13.6	6	9.1	6	9.1
Further Study	10	15.2	2	3.0	2	3.0
Language Study	0	0	1	1.5	0	0
GCSE	3	4.5	0	0	0	0
Other	1	1.5	0	0	2	3.0

A breakdown of the patterns of responses was also generated. It appears from Table 35.2 that learners were more likely to mention a particular FS image in isolation in an interview than they were in the survey. The fact that only 10 of the total 37 students who mentioned *Travel* in their interview mentioned this across both surveys too is an area for exploration.

One possible explanation for this is the 'group effect' of the interview. By way of an example, one participant could mention a possible or potential future self-image, which then leads subsequent interviewees to make reference to this. This does not necessarily

mean that these subsequent references are disingenuous, although the high occurrence of 'if' statements suggests that some instances may indeed have been. Rather, the mentioning of a possible future self-concept could have triggered this thought in other interviewees, to whom the use of Arabic for travel, for example, may not have previously occurred.

Table 35.2. Occurrences of AFS references across instruments				
	Interview only	Survey Only	Interview + 1 survey	All 3
Travel	15	4	9	10
Family	2	1	1	4
Religion	10	2	5	4
Job	7	5	2	5
Skill	8	10	1	0
Further Study	8	1	1	1
Language Study	0	1	0	0
GCSE	3	0	0	0
Other	1	2	0	0

It is difficult to examine the existence, or extent of group effects within interviews. One approach is to examine the nesting of certain codes within groups to see whether or not there were high levels of agreement on certain topics between students. Should this be the case for multiple items, this would call into question the reliability of the interview data. The occurrence of Arabic FS codes within interview groups was elicited from the data. Table 35.3 presents the percentage of learners in each group who mentioned different Arabic FS concepts.

Table 35.3. Interview Response Frequencies by Interview Group

	Classes 4 + 5							Class 6	Class 8			Classes 10 + 11				Class 12			Class 13		Class 16		Class 17				
Nodes	n=4	n=3	n=3	n=4	n=3	n=4	n=4	n=3	n=4	n=5	n=3	n=5	n=5	n=5	n=3	n=4	n=4	n=2	n=3	n=3	n=2	n=4	n=4	n=4	n=4	n=3	Total (n=95)
Travel	100%	67%	100%	75%			25%				33%	60%	40%	40%		25%	75%		100%		50%	100%			50%	67%	39%
Family	25%		33%	25%			100%	67%				20%								33%		25%					13%
Religion	25%	33%	67%	50%	33%	75%	75%	67%		20%	33%	20%	20%	20%					33%			25%	50%	25%	25%	100%	31%
Job			33%	50%	100%	50%		33%					20%			50%		100%		100%	50%	50%		25%	50%	33%	25%
Skill							25%					60%				75%	50%			33%		25%	50%	50%			16%
Communication		33%		25%		25%		67%			33%	20%	20%			50%				67%					25%		14%
Just GCSE		67%		25%				33%			33%			20%							50%	25%					8%
Other																50%									67%		4%
No Arabic		33%						33%	100%	100%	67%		40%	20%	67%	50%	25%						25%	75%			26%
'If'		67%	100%	25%	33%	25%	25%		25%	40%	67%	100%	40%		33%	50%	50%	50%	67%	67%	50%	25%					38%

In 11 of the 26 group interviews, there was repeated mention by all participants of specific Arabic FS images. This may explain, to some extent the higher frequency of references to certain FS images in interviews when compared to surveys which were completed independently. However, in all but 2 groups (Class 8 Group 1; Class 12 Group 3), there were also mentions of other FS images without the same agreement, implying at the very least that where agreement is found within interviews, this does not dictate the exclusion of alternative opinions.

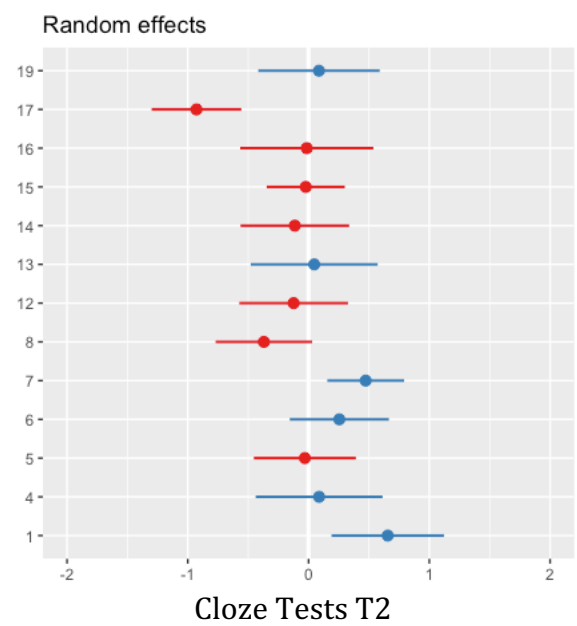
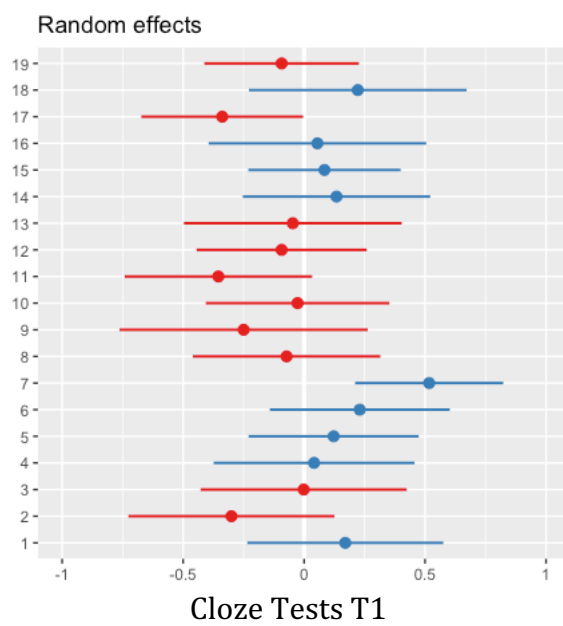
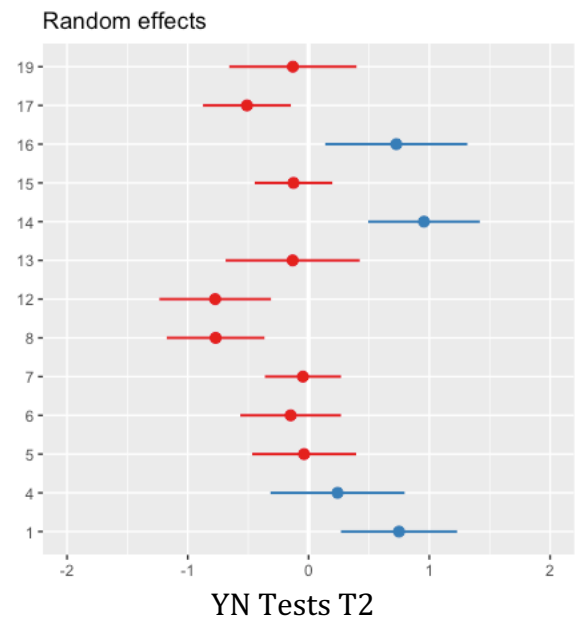
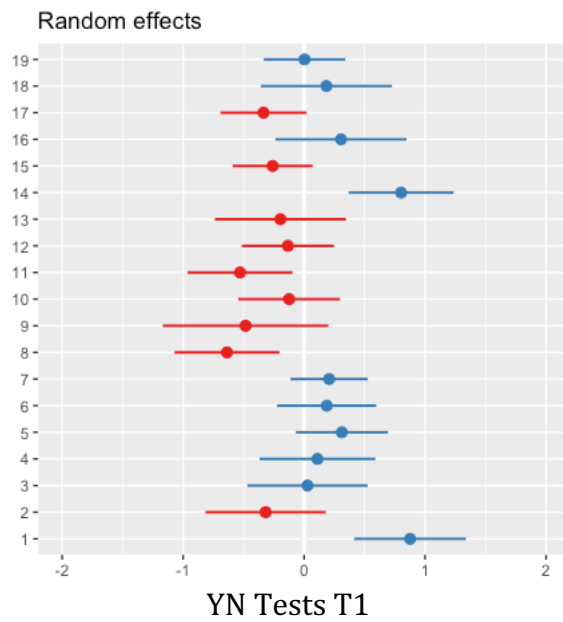
While this does not seem to be an area of concern in this particular study in the context of Arabic FS images, the nesting of replies within interview groups is, at any rate, an area needing further investigation.

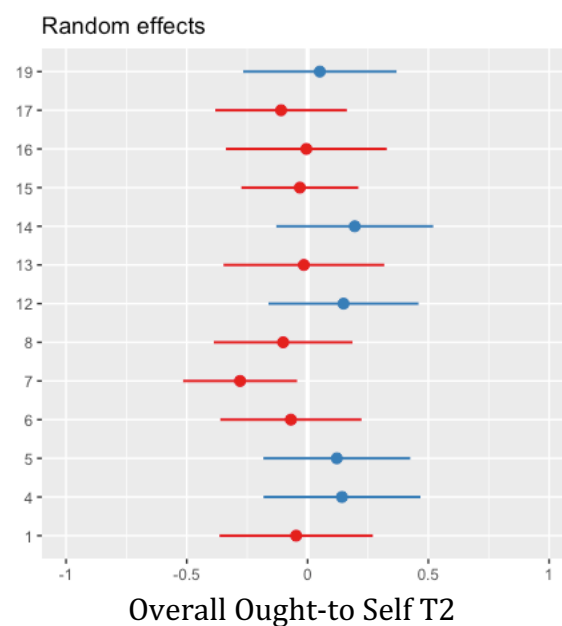
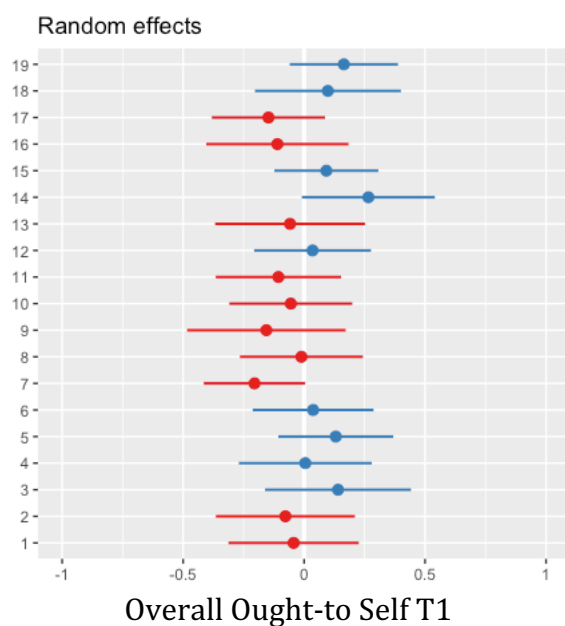
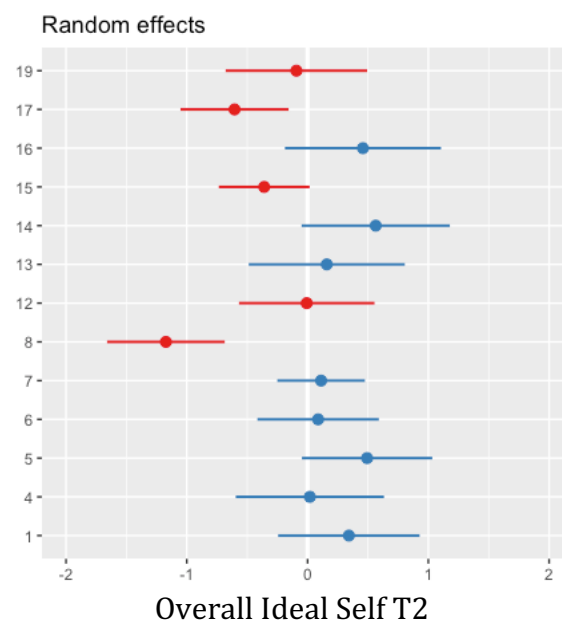
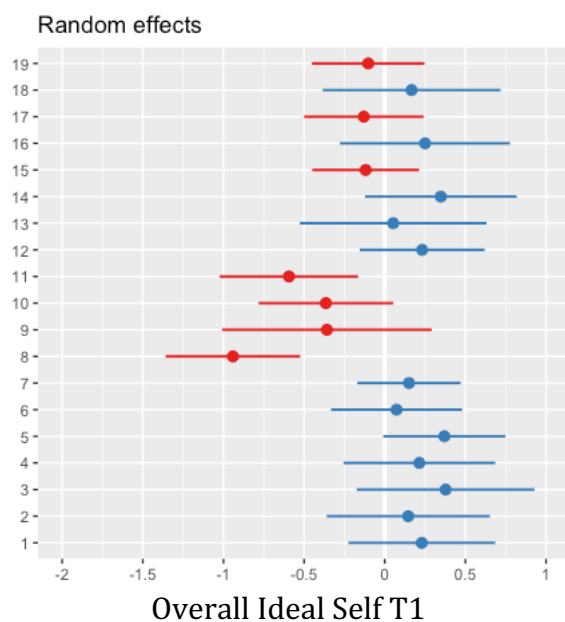
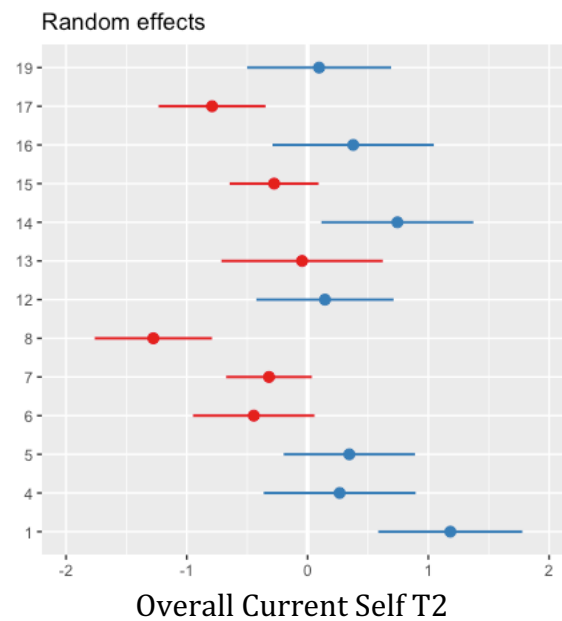
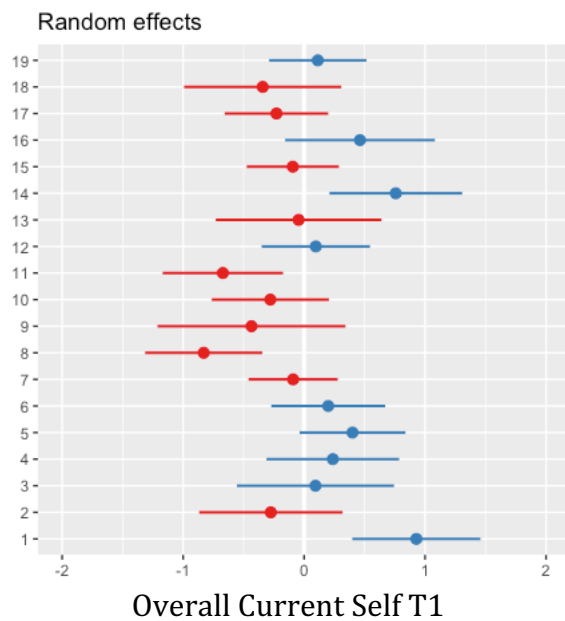
What was immediately interesting was the predominance of certain Arabic FS images within school groups. Appendix C.33 presents the case frequencies and percentages per school per code. Within Class 8, 91% of students expressed the absence of a FS image related to Arabic. Incidentally, this accounts for one of the interview groups in which there was 100% agreement amongst participants on a single code. This alludes to the fact that any nesting is taking place at the school level, rather than the interview group. This high propensity of participants in Class 4 + 5 and Class 16 to mention *Travel*, for example, or *Religion*, which was mentioned in every interview in Class 4 + 5, also suggests a school/classroom-level degree of nesting. This point is reaffirmed by references within interviews referring to school level factors, predominantly teachers and teaching approaches which may or may not impact learners' Ideal Selves, or at the very least their Current Self concepts.

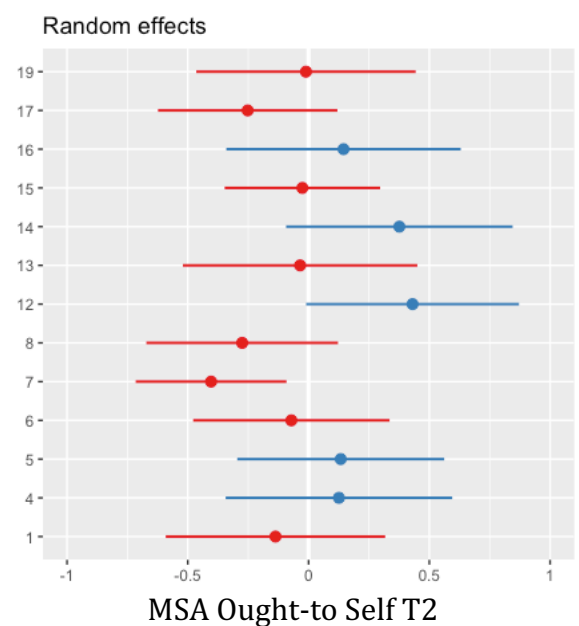
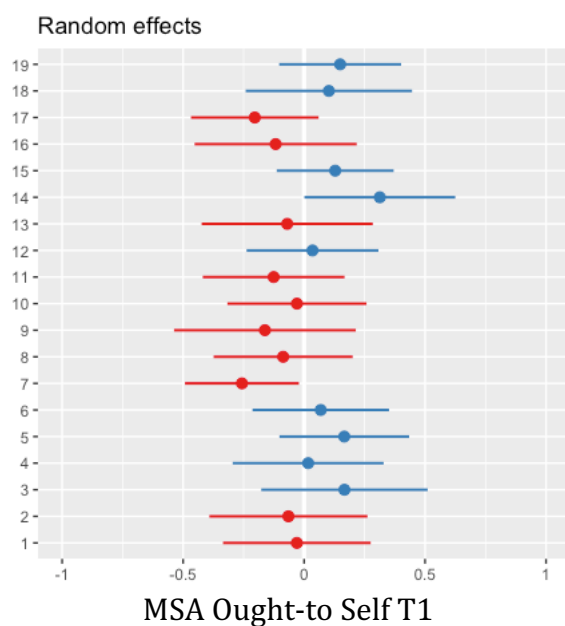
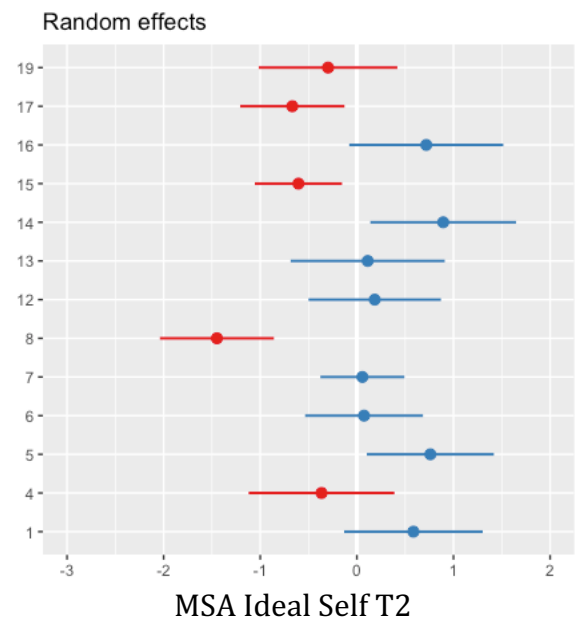
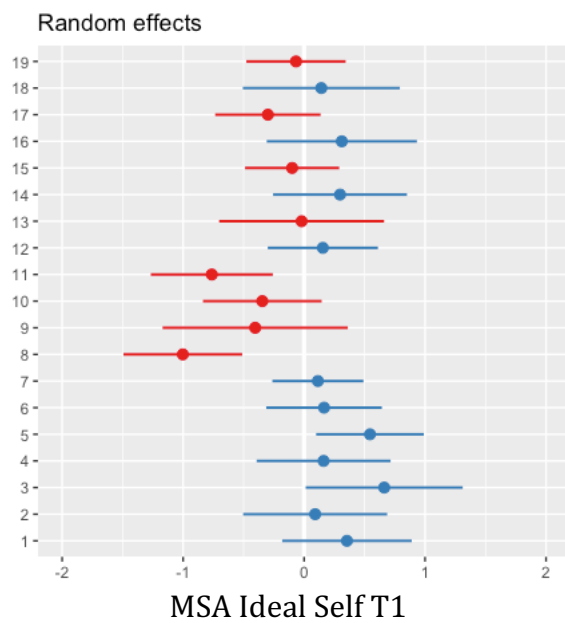
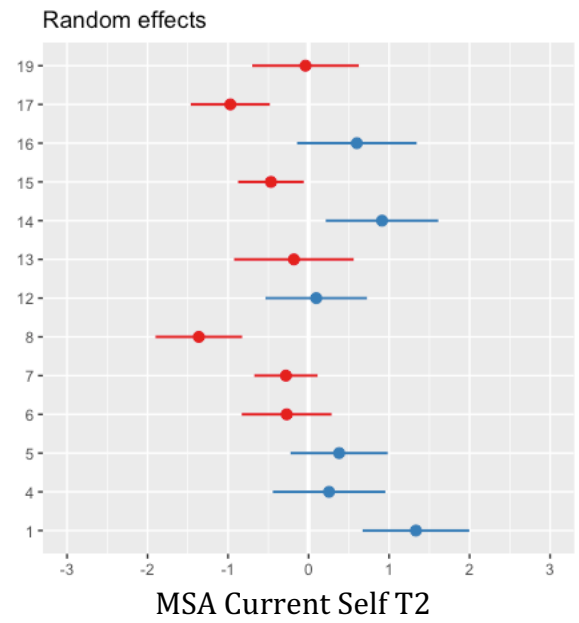
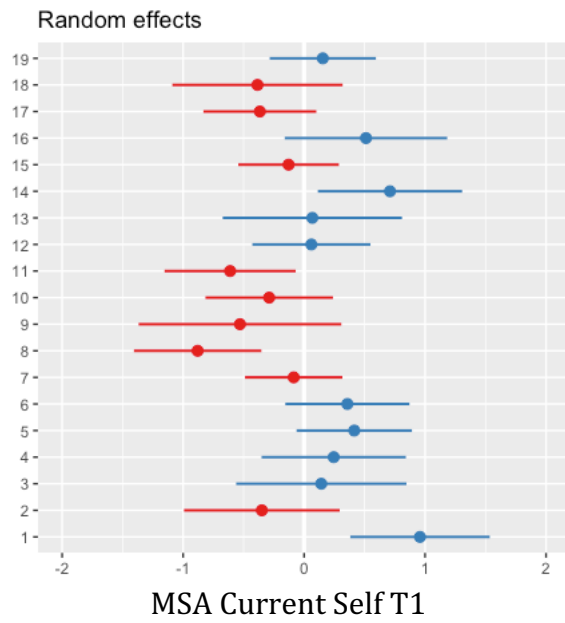
The school level clustering also sheds light on the gender differences that emerged in regard to Arabic FS images. The higher levels of *No Arabic* are clearly at least partially explained by the fact that 11 out of 12 students in one school (Class 8), an all-girls school, mentioned this in their interviews. Given the hypothesized link between Current Self and Ideal Self, this may also explain why there were higher levels of despondency amongst female students, if at least one school appears to be highly disenfranchised from any motivation to learn the language. This school-level clustering should be taken

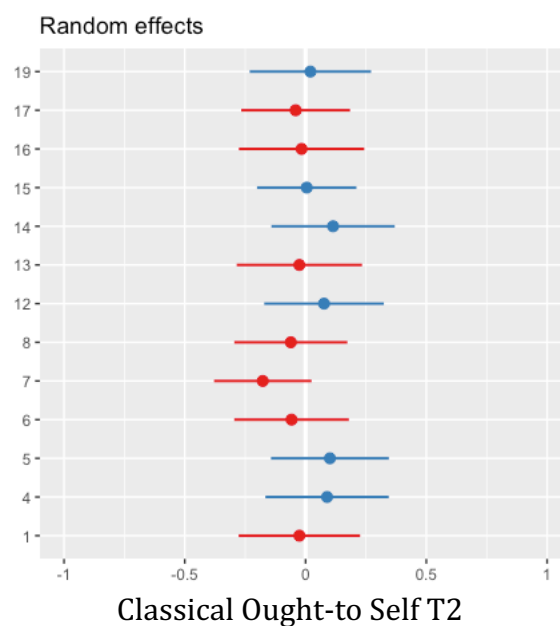
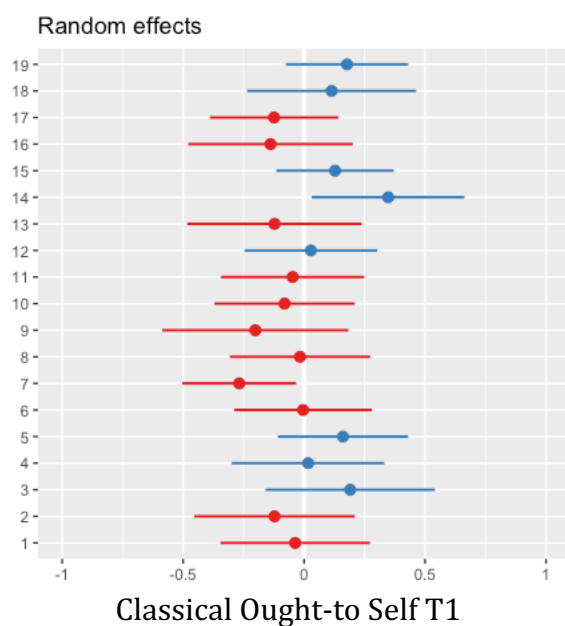
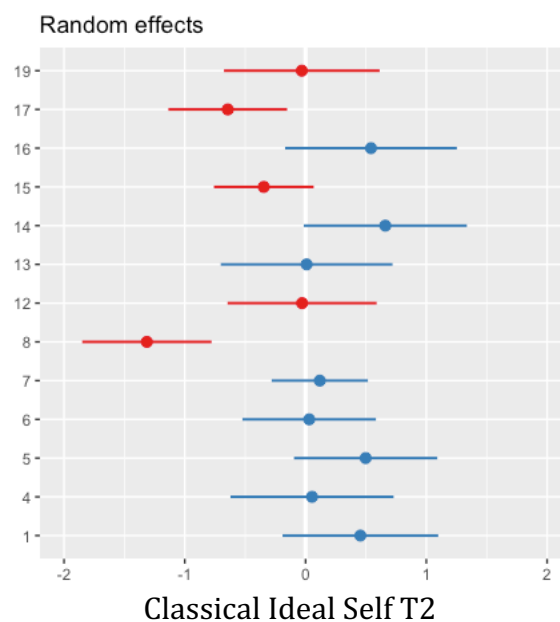
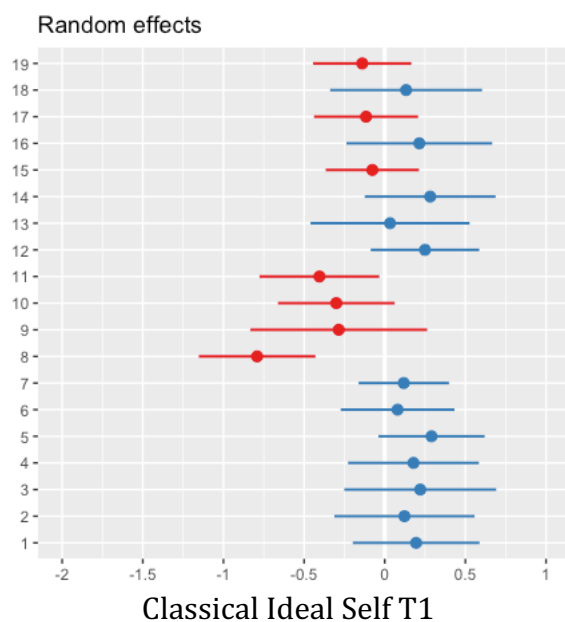
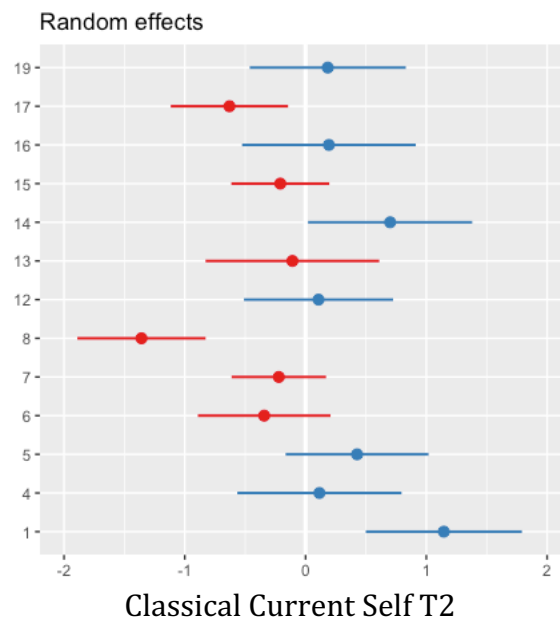
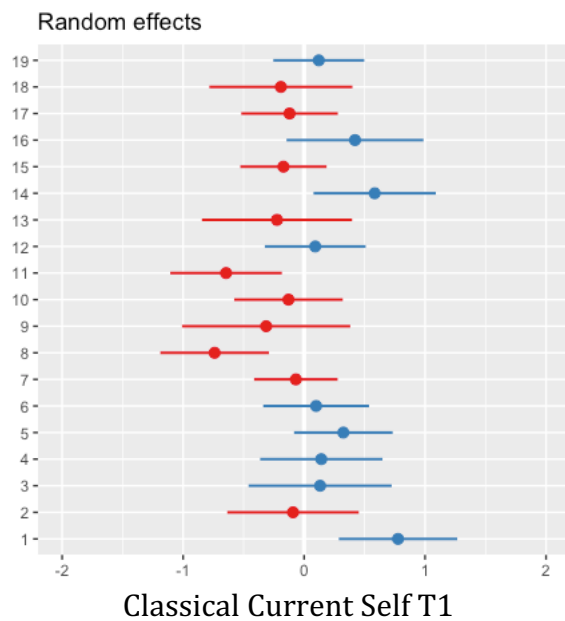
into account, where possible, when examining the relationship between motivational constructs over time in subsequent research questions.

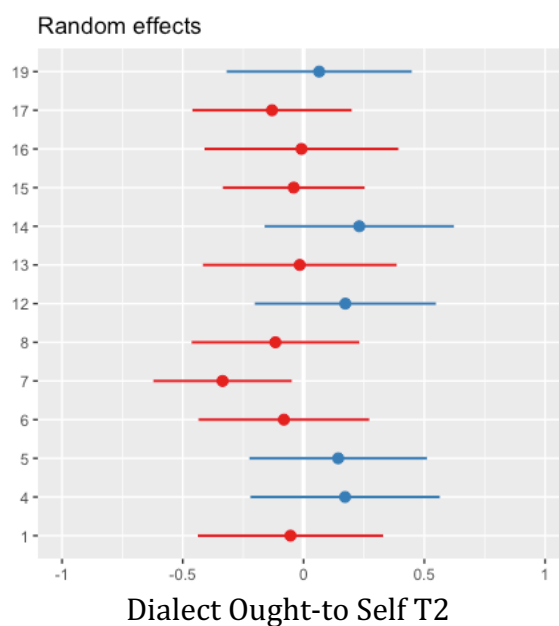
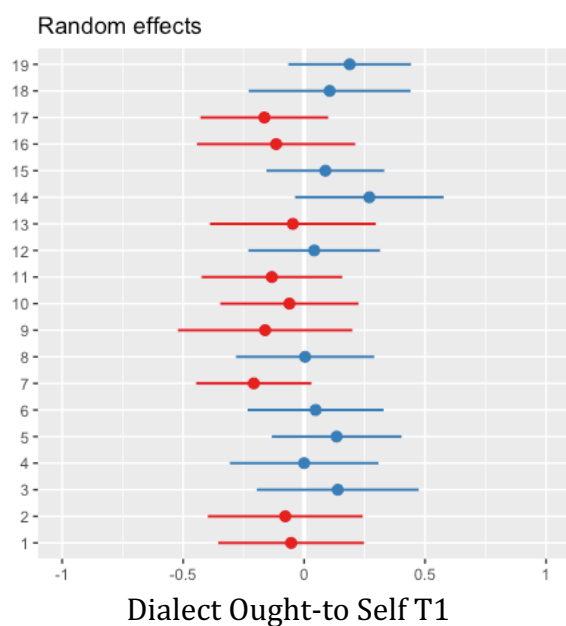
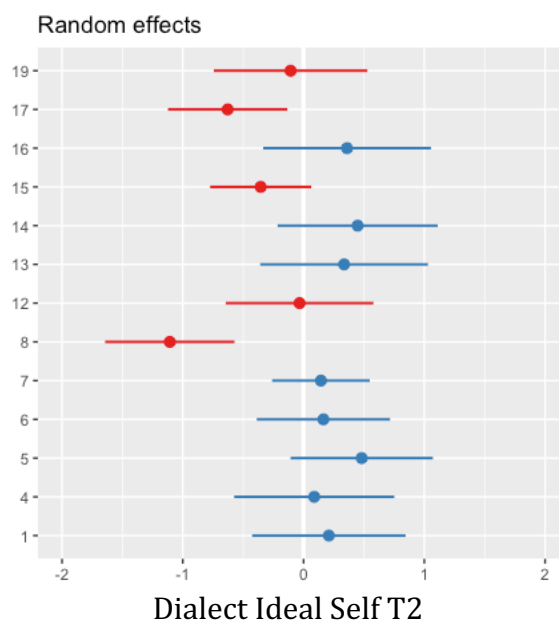
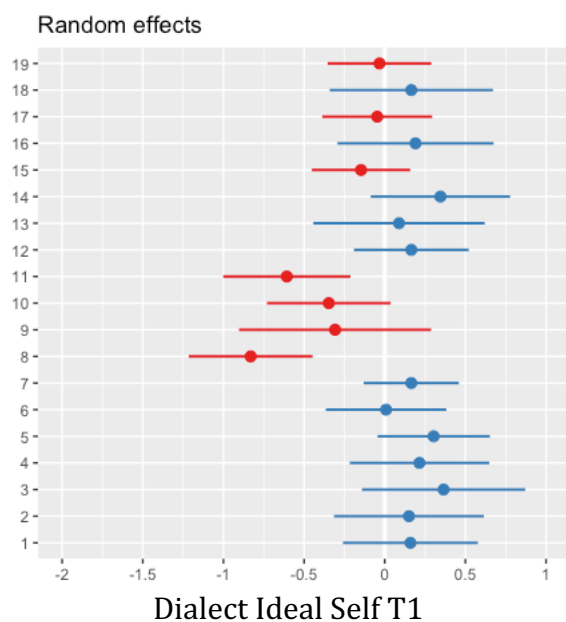
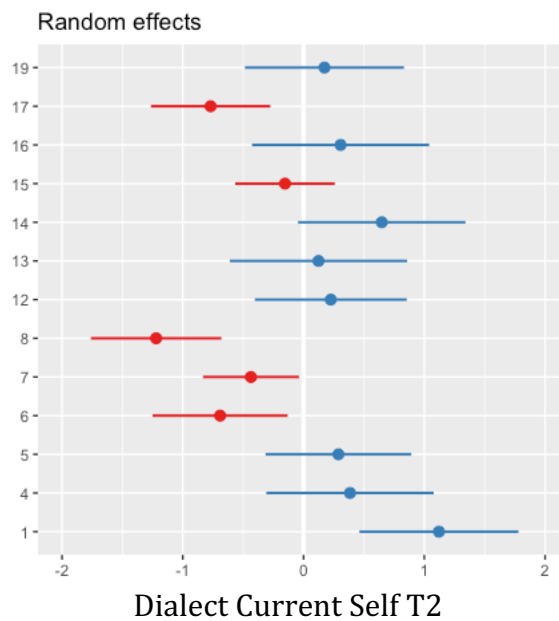
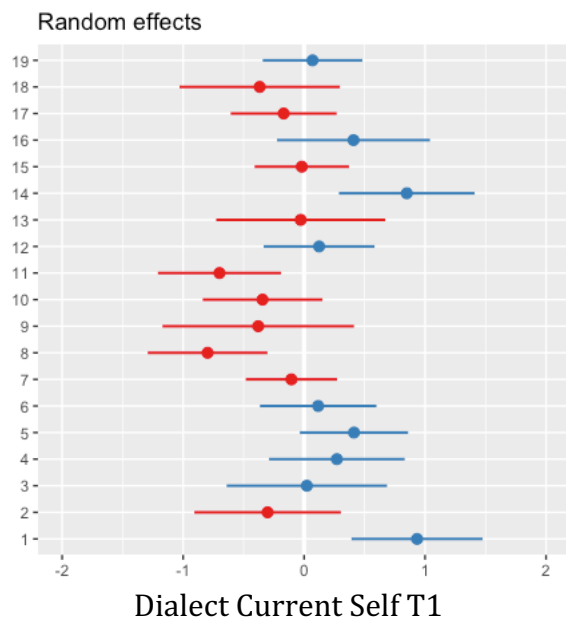
C.42. Caterpillar Plots for Self-Concept and Proficiency Test Variables











C.43. Descriptive Statistics for Self-Discrepancy Sum Scores

Self-Discrepancy	N	Range	Min.	Max.	Mean	SD	Skew.	SE	z-score	Kurtosis	SE	z-score
MSA Discrepancy T1	264	8.000	-3.000	5.000	0.886	1.299	0.246	0.150	1.640	0.190	0.299	0.635
Classical Discrepancy T1	264	7.333	-3.333	4.000	0.662	1.266	0.066	0.150	0.440	0.471	0.299	1.575
Dialect Discrepancy T1	265	7.667	-3.000	4.667	1.006	1.336	0.285	0.150	1.900	0.051	0.298	0.171
Overall Discrepancy T1	265	6.333	-2.222	4.111	0.845	1.159	0.322	0.150	2.147	-0.053	0.298	-0.178
MSA Discrepancy T2	174	7.333	-3.000	4.333	0.705	1.408	0.310	0.184	1.685	-0.201	0.366	-0.549
Classical Discrepancy T2	174	6.667	-3.000	3.667	0.722	1.270	0.050	0.184	0.272	-0.181	0.366	-0.495
Dialect Discrepancy T2	174	5.667	-1.667	4.000	1.412	0.978	0.175	0.184	0.951	0.210	0.366	0.574
Overall Discrepancy T2	174	4.222	-1.333	2.889	0.946	0.742	0.007	0.184	0.038	-0.207	0.366	-0.566

C.44. Correlation Matrices for IS Plausibility and Self-Discrepancy

T1		Spearman's rho	MSA CS-IS Disc	CI CS-IS Disc	D CS-IS Disc
MSA Plausibility	Corr. Coeff.		0.03	-0.049	-0.11
	Sig.		0.634	0.439	0.083
	N		250	250	250
CI Plausibility	Corr. Coeff.		-.135*	-.171**	-.175**
	Sig.		0.033	0.007	0.005
	N		250	250	250
D Plausibility	Corr. Coeff.		-0.101	-0.09	-.158*
	Sig.		0.111	0.156	0.013
	N		250	250	250

T2		Spearman's rho	MSA CS-IS Disc	CI CS-IS Disc	D CS-IS Disc
MSA Plausibility	Corr. Coeff.		0.006	-0.048	-0.111
	Sig.		0.945	0.555	0.173
	N		152	152	152
CI Plausibility	Corr. Coeff.		-0.037	-0.072	-0.015
	Sig.		0.652	0.379	0.851
	N		152	152	152
D Plausibility	Corr. Coeff.		-.178*	-0.145	-.190*
	Sig.		0.028	0.075	0.019
	N		152	152	152

Significance indicators:

* $p \leq .05$

** $p \leq .01$

C.45. Self-Discrepancy Path Model Output

Predicted Variable			Predictor Variable	Estimate	St. Error	z-value	p-value
Regressions							
SR Proficiency T2	~		Self-Discrepancy T1	0.007	0.026	0.273	0.785
SR Proficiency T2	~		Self-Rated Proficiency T1	0.558	0.065	8.547	0.000
SR Proficiency T2	~		Cloze Total T1	-0.024	0.033	-0.719	0.472
SR Proficiency T2	~		YN Total T1	0.100	0.040	2.473	0.013
YN Total T2	~		Self-Discrepancy T1	0.066	0.071	0.928	0.353
YN Total T2	~		YN Total T1	0.554	0.106	5.216	0.000
YN Total T2	~		Self-Rated Proficiency T1	0.281	0.146	1.923	0.054
Cloze Total T2	~		Self-Discrepancy T1	0.002	0.044	0.043	0.966
Cloze Total T2	~		Cloze Total T1	0.604	0.085	7.138	0.000
Cloze Total T2	~		Self-Rated Proficiency T1	0.306	0.149	2.050	0.040
Self-Discrepancy T2	~		Self-Discrepancy T1	0.631	0.067	9.461	0.000
Self-Discrepancy T2	~		Self-Rated Proficiency T1	-0.166	0.157	-1.061	0.289
Self-Discrepancy T2	~		Cloze Total T1	0.058	0.076	0.762	0.446
Self-Discrepancy T2	~		YN Total T1	-0.007	0.099	-0.073	0.942
Covariances							
Self-Discrepancy T1	~~		Self-Rated Proficiency T1	-0.177	0.048	-3.682	0.000
Self-Discrepancy T1	~~		Cloze Total T1	-0.089	0.089	-1.004	0.315
Self-Discrepancy T1	~~		YN Total T1	-0.320	0.094	-3.413	0.001
Self-Rated Proficiency T1	~~		Cloze Total T1	0.247	0.049	5.025	0.000
Self-Rated Proficiency T1	~~		YN Total T1	0.363	0.048	7.568	0.000
Cloze Total T1	~~		YN Total T1	0.505	0.085	5.957	0.000
SR Proficiency T2	~~		SR Proficiency T2	0.105	0.013	7.844	0.000
YN Total T2	~~		YN Total T2	0.550	0.104	5.308	0.000
Cloze Total T2	~~		Cloze Total T2	0.491	0.078	6.309	0.000
Self-Discrepancy T2	~~		Self-Discrepancy T2	0.590	0.071	8.325	0.000
Self-Discrepancy T1	~~		Self-Discrepancy T1	1.133	0.142	7.991	0.000
Self-Rated Proficiency T1	~~		Self-Rated Proficiency T1	0.289	0.033	8.675	0.000
Cloze Total T1	~~		Cloze Total T1	0.863	0.076	11.400	0.000
YN Total T1	~~		YN Total T1	0.867	0.095	9.169	0.000
SR Proficiency T2	~~		YN Total T2	0.058	0.021	2.820	0.005
SR Proficiency T2	~~		Cloze Total T2	-0.002	0.020	-0.098	0.922
SR Proficiency T2	~~		Self-Discrepancy T2	-0.027	0.022	-1.203	0.229
YN Total T2	~~		Cloze Total T2	0.211	0.063	3.352	0.001
YN Total T2	~~		Self-Discrepancy T2	-0.046	0.041	-1.134	0.257
Cloze Total T2	~~		Self-Discrepancy T2	0.012	0.042	0.290	0.772