

Language identification in Iran

by

Parisa Switzer (Sabethematabadi)

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Applied Linguistics and Discourse Studies

Carleton University

Ottawa, Ontario

©2021, Parisa Switzer (Sabethematabadi)

Abstract

For many years scholars have endeavoured to define what a language is and address how it is different from other types of language varieties such as dialect and accent. Some studies identify language varieties based on their structural similarity, focusing on the lexical similarity, and the phonological, morphological and syntactic properties. Others adopt a functional approach and focus on the degree to which speakers of language varieties "...can understand each other based on knowledge of their own language variety without needing to learn the other language variety, at a functional level" (Eberhard et al., 2021, p.1). Some studies address language identification issues following an ethnographic approach. These studies investigate language as a communicative behavior used by members of a particular culture (Hymes, 1989). Finally, studies which adopt a folk linguistic approach to language underline the linguistic perspectives of folk (non-specialists). Available studies in the field of Iranian linguistics adopt a predominantly structural approach to identify different types of language varieties. Given this imbalance in the literature, this dissertation chronicles the development of a measurement tool in the context of an exploratory mixed methods research project. Based on available literature, initial qualitative study results and an expert panel's views, I developed a sociolinguistic questionnaire. Following rigorous steps in questionnaire validation (Hinkin, 1998), I validated this questionnaire with 5 components and 21 statements. To treat folk perspectives as meaningful in their own right alongside the perspectives of scholars, this questionnaire investigated the linguistic perspectives of participants from a cross-section of speech communities in Iran. The questionnaire further identified the parameters associated with these perspectives in defining and differentiating different types of language

varieties. Results indicated that parameters including language variety status, extent of mutual intelligibility and shared words between language varieties, and standardized writing systems were the parameters which statistically significantly influenced Iranian folk perspectives. Perspectives regarding the relevance of urban vs. rural geography and number of speakers were heterogeneous. Results further showed that in language identification, the participants' education, prior formal knowledge of linguistics, and mother tongue predicted 53% of the variation in their linguistic perspectives among the sample population.

Acknowledgements

Writing this dissertation has been an invaluable and challenging learning journey. I am indebted to many individuals who supported this effort and without whom this journey would not have been possible. It has been a great privilege to study under the supervision of Dr. Erik Anonby, Dr. Trudy O'Brien and Dr. Jaffer Sheyholislami. I could not have asked for a more caring and understanding supervisory team. Not only did they offer their unconditional academic support, they also stood by my side in the most challenging moments of my personal life.

Trudy, you are the definition of irreplaceable. Like a mother, you were a light in my darkest days and a shoulder to cry on in the toughest of times. You always celebrated every small victory with me. Thank you for making me and my work a priority and for patiently reading and commenting on every document that I ever submitted to you. Erik, your approachable demeanor was the key to our productive, open and honest conversations all these years. Thank you for your thorough teaching approach, in-depth explanations of challenging concepts and for having long talks with me to make sure that I could figure out my next steps. Ostad, thank you for deepening my understanding of the significance of the social aspects of my work and helping me develop my academic writing skills. I would also like to thank Dr. Mako Hirotsu and Dr. Tamara Sorenson Duncan for improving the quality of my dissertation with their continued guidance during data collection and analysis procedure.

I wish to express my heartfelt appreciation to my family members for their consideration and patience. Casey, the thought of all that you went through in the past five years makes me believe in the fairytale version of love. It is beyond me how loving, caring, and patient you have been all

these years. Mozhdeh and Amir, you took care of me and gave me hope in days when I was stuck and saw no way out of the dark hole I was inside. Maman, I love you. Thank you for giving me a solid foundation and for being proud of me in all moments of my life.

I would like to dedicate this dissertation to Jamshid Sabethematabadi, my father, my hero and the best math professor who taught me that staying on the beaten path is not always the best way towards success in life. I wish I were still that seven-year-old girl walking on the edge of the sidewalk and you, that calm dad who patiently held my hand so that I could keep my balance and continue walking as if I were completing a critical task. No wonder I am still on the edge of all sidewalks of my life, Baba. I walk on the edge because you taught me that fear is only a barrier to overcome. I walk on the edge because I know even if I fall, you would give me the same look that you always did every time that I felt like a failure, the look that meant “take a couple of minutes but only a couple of minutes because you should dust it off and keep going”. Even now, I feel like I am still walking on the edge because I chose to be a folk linguist, a career which may not be inspiring to all, but to me is worth pursuing, even if it means walking on the edge of the normal path. So, I thank you Baba, with all my heart. I hope you would be proud.



Table of Contents

Abstract	II
Acknowledgements.....	IV
Table of Contents	VI
List of Tables	XII
List of Figures	XIX
List of Appendices	XX
Abbreviations used in this dissertation	XXI
Glossary of Farsi terms	XXIII
1. Introduction	1
1.1. Language identification.....	2
1.2. A mixed methods approach to language identification.....	7
1.2.1. Phase one: Qualitative strand- semi-structured interviews.....	8
1.2.2. Phase two: Quantitative strand- development of a measurement tool.....	9
1.2.3. Phase three: Quantitative strand- administration of the questionnaire	9
1.3. Hybrid structure of this dissertation	11
2. Hybrid thesis: Part one	14
2.1. Literature Review.....	15
2.1.1. Approaches to language identification.....	16

2.1.2.	The significance of folk research.....	19
2.1.3.	Who folk are	30
2.1.4.	Aspects related to language studies in Iran.....	33
2.2.	Conceptual Framework: Theoretical triangulation	45
2.3.	Methodology	48
2.3.1.	Participants.....	48
2.3.1.1.	Phase one: Participants in semi-structured interviews	49
2.3.1.2.	Phase two: Participants in three pilot iterations	52
2.3.1.3.	Phase two: Participants in questionnaire validation process.....	57
2.3.1.4.	Phase three: Participants in administration of the questionnaire.....	62
2.3.2.	Overview of data collection and analysis	64
3.	Hybrid thesis: Part two (Phase one)	67
3.1.	Phase one- Step one: Ethics and assembly of an expert panel.....	69
3.2.	Phase one- Step two: Qualitative exploration of folk linguistic perspectives.....	70
3.2.1.	Thematizing	72
3.2.2.	Designing and interviewing	72
3.2.3.	Transcribing and analyzing: Eclectic approach to coding	75
3.2.4.	Themes generated based on semi-structured interviews.....	79
3.2.4.1.	Geography (urban vs. rural)	79
3.2.4.2.	Number of speakers.....	80
3.2.4.3.	Status of language varieties.....	80

3.2.4.4. Mutual intelligibility	81
3.2.4.5. Writing systems and extent of shared words.....	84
3.2.4.6. Exposure to language varieties.....	85
3.2.4.7. Setting and purpose of communication.....	86
3.2.4.8. Demographic information	86
3.2.4.9. Relationships	87
3.2.4.10. Diglossia and bilingualism	88
3.2.4.11. Language and identity	89
3.2.4.12. Religious affiliation.....	90
4. Hybrid thesis: Part two (Phase two)	92
4.1. Phase two- Step three: Components of the questionnaire and item generation	94
4.1.1. C1-GE (Geography).....	95
4.1.2. C2-NS (Number of Speakers).....	95
4.1.3. C3-ST (Status)	96
4.1.4. C4-MI (Mutual Intelligibility)	96
4.1.5. C5-SW (Shared Words)	96
4.1.6. C6-WS (Writing System).....	98
4.1.7. C7- EX (Exposure)	98
4.2. Phase two- Step four: Rating scales of the items	101
4.3. Phase two- Step five: Developing the preliminary demographic questions.....	102
4.4. Phase two- Step six: Item checking via expert review.....	102

4.5.	Phase two- Step seven: First pilot	106
4.5.1.	C1-GE (Geography).....	107
4.5.2.	C2-NS (Number of Speakers).....	108
4.5.3.	C3-ST (Status)	109
4.5.4.	C4-MI (Mutual Intelligibility)	110
4.5.5.	C5-SW (Shared Words)	111
4.5.6.	C6-WS (Writing System).....	112
4.6.	Phase two- Step eight: Integration of data via methodological triangulation	113
4.7.	Phase two- Step nine: Pilots 2 and 3	120
4.8.	Phase two- Step ten: Validation	130
5.	Hybrid thesis: Part two (Phase three)	139
5.1.	Phase three- Step eleven: Administration of the validated questionnaire.....	140
5.1.1.	C1-GN (Geography and Number of speakers)	141
5.1.1.1.	Overall results	141
5.1.1.2.	C1-GN (Geography and Number of speakers): Mother tongue.....	143
5.1.1.3.	C1-GN (Geography and Number of speakers): Formal linguistic knowledge..	145
5.1.1.4.	C1-GN (Geography and Number of speakers): Education	147
5.1.1.5.	C1-GN (Geography and Number of speakers): Age.....	148
5.1.1.6.	C1-GN (Geography and Number of speakers): Gender.....	150
5.1.2.	C2-ST (Status)	151
5.1.2.1.	Overall results	151

5.1.2.2.	C2-ST (Status): Mother tongue.....	152
5.1.2.3.	C2-ST (Status): Formal linguistic knowledge.....	153
5.1.2.4.	C2-ST (Status): Education	154
5.1.2.5.	C2-ST (Status): Age.....	155
5.1.2.6.	C2-ST (Status): Gender.....	156
5.1.3.	C3-MI (Mutual Intelligibility)	157
5.1.3.1.	Overall results	157
5.1.3.2.	C3-MI (Mutual Intelligibility): Mother tongue.....	159
5.1.3.3.	C3-MI (Mutual Intelligibility): Formal linguistic knowledge.....	160
5.1.3.4.	C3-MI (Mutual Intelligibility): Education	161
5.1.3.5.	C3-MI (Mutual Intelligibility): Age.....	162
5.1.3.6.	C3-MI (Mutual Intelligibility): Gender.....	163
5.1.4.	C4-SW (Shared Words).....	164
5.1.4.1.	Overall results	164
5.1.4.2.	C4-SW (Shared Words): Mother tongue.....	165
5.1.4.3.	C4-SW (Shared Words): Formal linguistic knowledge	166
5.1.4.4.	C4-SW (Shared Words): Education	167
5.1.4.5.	C4-SW (Shared Words): Age.....	168
5.1.4.6.	C4-SW (Shared Words): Gender.....	169
5.1.5.	C5-WS (Writing System).....	170
5.1.5.1.	Overall results	170

5.1.5.2. C5-WS (Writing System): Mother tongue	172
5.1.5.3. C5-WS (Writing System): Formal linguistics knowledge	172
5.1.5.4. C5-WS (Writing System): Education.....	173
5.1.5.5. C5-WS (Writing System): Age	174
5.1.5.6. C5-WS (Writing System): Gender	175
5.1.6. Multiple regression analysis	177
6. Discussion and Conclusion.....	186
6.1. Limitations and future research.....	194
6.2. Final reflections.....	198
Appendices.....	200
Appendix A	200
Appendix B	205
Appendix C	213
References.....	223

List of Tables

Table 1. Questionnaire developmental steps in three phases of a mixed methods study	8
Table 2. PhD project calendar- Exploratory mixed methods research project (2017-2021)	10
Table 3. Participants’ demographic information in the qualitative phase	49
Table 4. Participants’ demographic information in the first pilot iteration	52
Table 5. Participants’ demographic information in the second pilot iteration.....	53
Table 6. Speech community members in the second pilot iteration	54
Table 7. Participants’ demographic information in the third pilot iteration	56
Table 8. Speech community members in the third pilot iteration.....	57
Table 9. Participants’ demographic information during the Exploratory Factor Analysis (EFA)	57
Table 10. Speech community members during the Exploratory Factor Analysis (EFA)	60
Table 11. Participants’ demographic information during Confirmatory Factor Analysis (CFA).	60
Table 12. Speech community members during Confirmatory Factor Analysis (CFA)	61
Table 13. Participants’ demographic information in questionnaire administration.....	62
Table 14. Speech community members in the administration of language identification questionnaire	64
Table 15. Stages of an interview investigation adapted from Kvale (1996).....	71
Table 16. Types of interview questions with examples	73
Table 17. Example question type following Kvale’s (1996) typology	74
Table 18. Extracted themes and sample quotes via an eclectic approach to coding.....	77
Table 19. Components of the questionnaire based on qualitative themes	94

Table 20. Pool of generated items based on the components	99
Table 21. Refined questionnaire based on experts' views	105
Table 22. C1-GE (Geography): Likert-scale responses	107
Table 23. C2-NS (Number of Speakers): Likert-scale responses	108
Table 24. C3-ST (Status): Likert-scale responses.....	109
Table 25. C4-MI (Mutual Intelligibility): Likert-scale responses	111
Table 26. C5-SW (Shared Words): Likert-scale responses	112
Table 27. C6-WS (Writing System): Likert-scale responses.....	113
Table 28. Language identification questionnaire with 21 items.....	121
Table 29. Second pilot iteration: C1-GE (Geography) and C2-NS (Number of Speakers) Likert-scale responses	125
Table 30. Third pilot iteration: C1-GE (Geography) and C2-NS (Number of Speakers) Likert-scale responses	125
Table 31. Second pilot iteration: C3-ST (Status) Likert-scale responses	127
Table 32. Third pilot iteration: C3-ST (Status) Likert-scale responses	127
Table 33. Second pilot iteration: C4-MI (Mutual Intelligibility) Likert-scale responses	128
Table 34. Third pilot iteration: C4-MI (Mutual Intelligibility) Likert-scale responses	128
Table 35. Second pilot iteration: C5-SW (Shared Words) Likert-scale responses.....	129
Table 36. Third pilot iteration: C5-SW (Shared Words) Likert-scale responses.....	129
Table 37. Second pilot iteration: C6-WS (Writing system) Likert-scale responses	129
Table 38. Third pilot iteration: C6-WS (Writing system) Likert-scale responses	129

Table 39. KMO and Bartlett’s test results	131
Table 40. Factor loadings based on Principal Component Analysis (PCA).....	133
Table 41. Reliability index based on the questionnaires collected during the Exploratory Factor Analysis (EFA)	134
Table 42. Model fit summary: CMIN	135
Table 43. Model fit summary: Comparative fit indices	135
Table 44. RMSEA.....	136
Table 45. Summary of maximum likelihood estimates to establish convergent and discriminant validities.....	137
Table 46. C1-GN (Geography and Number of speakers): Likert-scale responses	142
Table 47. Farsi, Kurdish, Turkic, and Mazandarani speakers’ perspectives in C1-GN (Geography and Number of speakers)	145
Table 48. Participants’ perspectives with respect to formal linguistic knowledge in C1-GN (Geography and Number of speakers)	146
Table 49. College, BA and MA/PhD participants’ perspectives regarding C1-GN (Geography and Number of speakers)	148
Table 50. Participants’ perspectives with respect to age in C1-GN (Geography and Number of speakers)	149
Table 51. Male and female perspectives in C1-GN (Geography and Number of speakers)	150
Table 52. C2-ST (Status): Likert-scale responses.....	152
Table 53. Farsi, Kurdish, Turkic, and Mazandarani speakers’ perspectives in C2-ST (Status).	153

Table 54. Participants’ perspectives with respect to formal linguistic knowledge in C2-ST (Status)	154
Table 55. College, BA and MA/PhD participants’ perspectives regarding C2-ST (Status).....	155
Table 56. Participants’ perspectives with respect to age in C2-ST (Status).....	156
Table 57. Male and female perspectives in C2-ST (Status).....	157
Table 58. C3-MI (Mutual Intelligibility): Likert-scale responses	158
Table 59. Farsi, Kurdish, Turkic, and Mazandarani speakers’ perspectives in C3-MI (Mutual Intelligibility)	160
Table 60. Participants’ perspectives with respect to formal linguistic knowledge in C3-MI (Mutual Intelligibility)	161
Table 61. College, BA and MA/PhD participants’ perspectives regarding C3-MI (Mutual Intelligibility)	162
Table 62. Participants’ perspectives with respect to age in C3-MI (Mutual Intelligibility).....	163
Table 63. Male and female perspectives in C3-MI (Mutual Intelligibility)	164
Table 64. C4-SW (Shared Words): Likert-scale responses	165
Table 65. Farsi, Kurdish, Turkic, and Mazandarani speakers’ perspectives in C4-SW (Shared Words).....	166
Table 66. Participants’ perspectives with respect to formal linguistic knowledge in C4-SW (Shared Words).....	167
Table 67. College, BA and MA/PhD participants’ perspectives regarding C4-SW (Shared Words).....	168

Table 68. Participants’ perspectives with respect to age in C4-SW (Shared Words).....	169
Table 69. Male and female perspectives in C4-SW (Shared Words)	169
Table 70. C5-WS (Writing System): Likert-scale responses.....	170
Table 71. Farsi, Kurdish, Turkic, and Mazandarani speakers’ perspectives in C5-WS (Writing System)	172
Table 72. Participants’ perspectives with respect to formal linguistic knowledge in C5-WS (Writing System).....	173
Table 73. College, BA and MA/PhD participants’ perspectives regarding C5-WS (Writing System)	174
Table 74. Participants’ perspectives with respect to age in C5-WS (Writing System)	175
Table 75. Male and female perspectives in C5-WS (Writing System).....	176
Table 76. Regression model summary	180
Table 77. Model fitness via ANOVA	181
Table 78. Coefficients, correlations and collinearity statistics in model 5 with five variables ..	182
Table 79. Farsi speech community members in the second pilot iteration.....	213
Table 80. Kurdish speech community members in the second pilot iteration	213
Table 81. Turkic speech community members in the second pilot iteration	213
Table 82. Mazandarani speech community members in the second pilot iteration.....	214
Table 83. Farsi speech community members in the third pilot iteration	215
Table 84. Kurdish speech community members in the third pilot iteration.....	215
Table 85. Turkic speech community members in the third pilot iteration.....	215

Table 86. Gilaki speech community members in the third pilot iteration	215
Table 87. Lori speech community members in the third pilot iteration.....	216
Table 88. Farsi speech community members during the Exploratory Factor Analysis (EFA)...	217
Table 89. Kurdish speech community members during the Exploratory Factor Analysis (EFA)	217
Table 90. Turkic speech community members during the Exploratory Factor Analysis (EFA)	217
Table 91. Mazandarani speech community members during the Exploratory Factor Analysis (EFA)	218
Table 92. Balochi speech community members during the Exploratory Factor Analysis (EFA)	218
Table 93. Lori speech community members during the Exploratory Factor Analysis (EFA)....	218
Table 94. Gilaki speech community members during the Exploratory Factor Analysis (EFA).	218
Table 95. Farsi speech community members during the Confirmatory Factor Analysis (CFA)	219
Table 96. Kurdish speech community members during the Confirmatory Factor Analysis (CFA)	219
Table 97. Turkic speech community members during the Confirmatory Factor Analysis (CFA)	220
Table 98. Mazandarani speech community members during the Confirmatory Factor Analysis (CFA)	220
Table 99. Farsi speech community members in the administration of the language identification questionnaire	221

Table 100. Kurdish speech community members in the administration of the language identification questionnaire.....	221
Table 101. Turkic speech community members in the administration of the language identification questionnaire.....	222
Table 102. Mazandarani speech community members in the administration of the language identification questionnaire.....	222

List of Figures

Figure 1. Scree plot with 5 components above eigenvalue 1.....	132
Figure 2. Final language identification questionnaire with 5 components identified during the EFA.....	138

List of Appendices

Appendix A.....	200
Appendix B.....	205
Appendix C.....	213

Abbreviations used in this dissertation

AILA	The International Association of Applied Linguistics
ALDS	Applied Linguistics and Discourse Studies
ALI	The Atlas of the Languages of Iran
AVE	Average Variance Extracted
CFA	Confirmatory Factor Analysis
CFI	Comparative Fix Index
CMIN/DF	Minimum Discrepancy per Degree of Freedom
CR	Composite Reliability
CUREB-A	Carleton University Research Ethics Board-A
EFA	Exploratory Factor Analysis
GFI	Goodness of Fit Index
IAUCTB	Islamic Azad University Central Tehran Branch
ICKS3	The Third Kurdish Studies Conference
IFI	Incremental Fit Index
IUST	Iran University of Science and Technology
KMO MSA	Kaiser Meyer Olkin Measures of Sampling Adequacy
LRR	The journal Language Related Research
PCA	Principal Component Analysis
RMSEA	Root Mean-Square Error of Approximation
RNI	Relative Non-centrality Index
SEM	Structural Equation Modelling
SRTTU	Shahid Rajaei Teacher Training University
TLI	Tucker-Lewis Index
Components of the language identification questionnaire during pilot iterations	
C1-GE	Component 1- Geography
C2-NS	Component 2- Number of Speakers
C3-ST	Component 3- Status

C4-MI	Component 4- Mutual Intelligibility
C5-SW	Component 5- Shared Words
C6-WS	Component 6- Writing System
Components of the validated language identification questionnaire	
C1-GN	Component 1- Geography and Number of Speakers
C2-ST	Component 2- Status
C3-MI	Component 3- Mutual Intelligibility
C4-SW	Component 4- Shared Words
C5-WS	Component 5- Writing System

Glossary of Farsi terms

Farsi	English
<i>Guneh</i>	Variety
<i>Guyeshshenāsi</i>	Dialectology
<i>Guyesh</i>	Common English equivalent: Dialect
<i>Lahjeh</i>	Common English equivalent: Accent
<i>Zabān</i>	Common English equivalent: Language

Preface

This dissertation is an original work by Parisa Switzer (Sabethematabadi¹). In the first phase of this exploratory mixed methods research project, I conducted 12 semi-structured interviews and analyzed the collected data. With Erik Anonby and Jaffer Sheyholislami, I presented the results of this project entitled “A comparative analysis of language classification taxonomies in Kurdish, Tehrani, and Abade’i speech communities” at the Third Kurdish Studies Conference (ICKS3) at the University of Exeter, UK. With my co-presenters’ permission, I included the findings of this qualitative study in Section 3 of this dissertation. During the second phase, I developed a language identification questionnaire (Appendix B). To establish reliability and validity of this questionnaire, I collected and analyzed 752 questionnaires. I led a paper with Erik Anonby as co-author, entitled “Language identification in Iran: A sociolinguistic model and questionnaire”. We submitted this paper to the journal *Language Related Research (LRR)*. With his permission, I elaborated on the findings of this study in Section 4. Finally, in the third phase, I collected 192 questionnaires to investigate folk linguistic perspectives. As lead author, in collaboration with Erik Anonby, I analyzed the collected data and elaborated on the results of this study in a paper entitled “An analysis of folk linguistic perspectives in Iran”. We submitted this paper to *AILA Review*. With his consent, I presented the results of this paper in Section 5. I have benefited from input on this thesis from the members of the thesis supervisory team and other committee members, and I have led and carried out all other aspects and components of this thesis.

¹ My conference presentation and current publications are under my maiden name.

1. Introduction

In this chapter, I briefly introduce the issues of language identification and scholars' approaches to address these issues (Section 1.1.). Some Iranian scholars (e.g., see Hassanpour 1992, 2000; Rezaei & Bahrami, 2019; Sheyholislami, 2012, 2015, 2019) underline the influence of sociocultural parameters in identifying language varieties. However, the majority of studies in Iranian linguistics identify language varieties via a structural approach. In these studies, there is a focus on the lexical similarity, and the phonological, morphological and syntactic properties of the language varieties under study. Nonetheless, in this section and throughout the dissertation, I underline the significance of taking a comprehensive and balanced approach which recognizes language varieties as social entities and identifies them with respect to folk (non-linguists) perspectives. Further, in answering the research questions of this study (see end of Section 1.1.), I draw attention to the importance of devising an appropriate and meaningful instrument by which to examine folk perspectives and determine the parameters associated with them. In addition, I explain the reasons why I adopt an exploratory sequential design to a mixed methods approach in identifying language varieties in Iran (Section 1.2.). Lastly, I elaborate on the hybrid structure of this dissertation which contains two main parts (Section 1.3.). These parts are representatives of chapters in a monograph dissertation as well as features of a manuscript dissertation. I also explain the challenges which influence my decision to follow a hybrid style instead of a manuscript (integrated articles) or a traditional monograph style dissertation.

1.1. Language identification

Although the discipline of linguistics has made significant efforts to define, differentiate, classify, and document languages, there is little consensus regarding the definition of language itself or what constitutes a language (Murphy, 1998). Language studies (Chomsky, 1972; Eberhard et al., 2021; Saville-Troike, 2003; Wolfram, 2017) have attempted to define and differentiate different types of language varieties “the issues of language identification” by delineating the ‘borderline’ between them. Distinguishing between types of language varieties based on linguistic labels such as ‘language’, ‘dialect’ and ‘accent’ seem to be self-evident to some of the scholars (Haugen, 1966; Perry, 1998; Windfuhr, 2009). Eberhard et al. (2021), however, colourfully describe attempts to identify language varieties via one approach or another as nailing jello to the wall. Such challenges are due to the fact that languages are in constant flux and associated with numerous parameters in specific contexts (Eberhard et al., 2021; Haig & Öpengin, 2014; Wolfram, 2017).

While Chomsky (1972) does not deny the influence of non-linguistic parameters on encoding and decoding speech, he detaches the domain of linguistics, since he considers these parameters as the concerns of social sciences. Chomsky’s ideal of a detached and pure linguistics, however, is not achievable in the real world. With the beginning of the ethnography of speaking in the 1950s and early 1960s (Hymes, 1974), scholars underscored the ‘meaningless’ enterprise of attempting to define types of language varieties in purely linguistic terms.

Scholars (e.g., see Bhatia et al., 2008; Gumperz & Hymes, 1972) underline language as an entity which cannot be separated from the essential and practical questions of ‘how’, ‘for what

purpose’, ‘where’, and ‘by whom’ it is used. Comprehensive studies rely on approaches where language varieties are identified and analyzed in relation to the characteristics of the societies where they are used (Chambers & Trudgill, 1998). In doing so, Hymes (1972) advocates comparative ethnographic studies to explore, understand, analyze, and compare languages systematically in speech communities. Nonetheless, to date, the greatest number of studies continue to determine differences between language varieties adopting a structural approach to language identification.

Despite the invaluable contributions of a structural approach to define and differentiate types of language varieties, language identification must “...ultimately depend on the discovery of which differences are recognized by members of the group as conveying social meaning of some kind” (Saville-Troike, 2003, p. 41). Adopting a strictly structural approach cannot satisfactorily shed light on perspectives among speech communities and in the wider society.

Language varieties are continuous entities which can be identified as a property of a temporal continuum, from past to present, and associated with individuals in the expanses of geographic and social space. Along these continuums, language varieties influence and are influenced by many parameters, both linguistic and non-linguistic. For example, among individuals, traces of such parameters are particularly evident in the speech of people who move away from home (Halliday, 1978). Considering the significance of these parameters, there is a need to address them, just as much, with respect to language identification in Iran (Sabethematabadi, et al., 2017).

To identify language varieties meaningfully, they must be investigated not only as the domain of linguists and sociolinguists, but also according to the perspective of “folk” – that is, non-linguists or non-experts – who use and speak these varieties (Hoenigswald, 1966; Preston, 1989). Recognizing the significance of how the community of non-linguists characterizes linguistic facts, Preston (1989, 1999a) strongly argues that perspectives of ordinary non-linguists constitute social scientific information and add to the body of scientific data in linguistics.

While scholars’ approaches to studying language varieties are valuable, increasing attention has been paid to folk perception of language varieties (Rezaei & Bahrami, 2019; Rezaei et al., 2017), folk linguistic perspectives regarding mapping language varieties (Preston, 1989, 1999a), status of language varieties and speech communities (Preston, 2016; Saviile-Troike, 2003), linguistic labels (Dabir-Moghaddam, 2008; Modarresi, 2014), and significance of community membership and sociocultural values (Fishman, 1965, 1972; Gumperz, 1964; Saviile-Troike, 2003) in language identification.

A number of linguistic studies have echoed persistent views in academia that what non-linguists say is uninteresting, unqualified, uninformed, dangerous, and arising from lack of education (e.g., see Bloomfield, 1944; Labov, 1972). However, Wilton and Stegu (2011) point out that focusing on folk linguistic data does not mean that only non-linguists hold the truth about language and its use, and insist that it is important to consider folk linguistic data as valuable information about language. The potential shortcomings of folk linguistic theory should not end in its exclusion, but rather should be considered a point of cross-fertilization for the ethnography of speaking, linguistics and sociolinguistics studies.

On significance of folk knowledge, Preston (1999a, p. xxiii) adds that “what people actually say not only provides ethnographic, conversational and other studies of language performance with raw data, but also by exposing how different groups say things differently, feeds (at least) historical linguistics, linguistic geography, and sociolinguistics”. Analysis of such raw data, in turn, opens a new door to the study of processes which govern what people say (Hoenigswald, 1966). Preston (1999a) further explains that these underlying processes are related to people’s language use and language attitude. He continues to say that how folk talk about ‘language’ is valuable even if there are major contrasts between academic and folk perspectives. Intervening successfully in such cases of contrast is of value specifically to those who work in applied fields such as language teaching and learning (Pasquale & Preston, 2013).

The majority of language identification research in Iran reveals that language varieties have been identified and classified from linguists’ perspectives (e.g., see Anonby, Taheri-Ardali, et al., 2015-2021; Borjian, 2008, 2013; Naghzguy-Kohan, 2016). These studies, which diverge in defining and differentiating different types of language varieties in Iran, paid attention to linguistic parameters and occasionally to political and sociocultural priorities as the additional influential parameters that researchers need to consider in establishing a proper understanding of Iranian linguistics and language varieties. The varied and opposing perspectives about fundamental terms such as ‘language’, ‘dialect’ and ‘accent’ are both concerning and confusing for linguistics students and researchers (Dabir-Moghaddam, 2008).

Moreover, each speech community in Iran might approach the issues of language identification differently and with reference to a selection of parameters that are not necessarily

shared among all those communities (Anonby, Hayes, et al., 2020; Modarresi, 2014; Sabethematabadi et al., 2017). This is partly due to the fact that Iran is a multilingual country, where many language varieties coexist due to cultural and economic influences, migration, as well as political, historical and religious struggles (e.g., see Hassanpour, 1992; Moradi, 2019).

Considering the respective challenges in defining and differentiating the language varieties and responding to the insufficiency of folk linguistic studies with a focus on language identification in Iran, this dissertation asks two main questions: First, *how do Iranian folk define and differentiate different types of language varieties in Iran?* And second, *what parameters influence Iranian folk perspectives about language varieties in Iran?*

In order to ascertain the answers to these questions, this dissertation acknowledges the urgency for a measurement tool which can address the issues of language identification with a large sample population and pave the way towards getting one step closer to better understanding the language situation in Iran. Folk data collected using this questionnaire will provide significant information about their attitude towards their own and others' language varieties. As Preston (1999a) asserts, completeness of the ethnography of language for any speech community relies on what non-linguists believe about language. After all, "why would we assume that linguists could not gain clues about language by listening to the linguistic comments of the folk?" (Preston, 2005, p. 145).

1.2. A mixed methods approach to language identification

A mixed methods approach to research is a methodology for conducting a study in social, health and behavioral sciences which involves collection and analysis of both qualitative and quantitative studies to find answers for the underlying issues (Creswell, 2015). Such an integrative approach to addressing the issues of language identification has helped overcome the inherent weaknesses of each of the qualitative or quantitative approaches alone. Although a qualitative orientation to research was much needed in establishing the underlying constructs in the components of the language identification questionnaire, a stand-alone qualitative research would not thoroughly address these issues. Considering that the qualitative approach to research was time-consuming and explored the issues of language identification among a limited number of perspectives, the results were neither representative of nor generalizable to a heterogeneous cross-section of speech communities in Iran. Furthermore, a quantitative approach to address these issues was particularly challenging due to the lack of a measurement tool to collect and analyze folk linguistic perspectives from speech communities in Iran.

In this dissertation, I adopt an exploratory sequential design to a mixed methods approach (Creswell, 2015). This three-phase design (qualitative-quantitative-quantitative) facilitated a preliminary exploration of folk linguistic perspectives during the qualitative phase. This phase, in turn, ensured that the measurement instrument was grounded in the actual experience of the participants and their contexts. Having determined the reliability and validity indices of the questionnaire, to investigate the participants' linguistic perspectives and whether their demographics influenced these perspectives, I administered it among 192 Iranian folk participants.

This design was sophisticated since conducting the three phases was time-consuming and creating a quantitative component based on qualitative results was extremely challenging. Nonetheless, as Creswell (2015) explains, it might be a good choice for research in countries where the measures of the Western world might have limited applicability and researchers first need to explore what measure will work in the setting. The following section summarizes the three phases of this exploratory sequential mixed methods research project.

1.2.1. Phase one: Qualitative strand- semi-structured interviews

In the first phase (see Table 1), I assembled a panel of experts and completed the ethics process through the Carleton University Research Ethics Board-A (CUREB-A). The panel's feedback was specifically helpful in generating questions for semi-structured interviews, generating statements in all sections of the questionnaire, establishing reliability and validity of the questionnaire, and data analysis and interpretation during the qualitative and quantitative phases. In Step two, I conducted semi-structured interviews with 12 participants to collect their perspectives about language varieties and identified the parameters that they found influential in addressing the issues of language identification. Analysis of collected qualitative data underlined multiple parameters meaningfully associated with linguistic perspectives (see Step two). These parameters helped develop 36 initial statements in the original version of the questionnaire (see Step three).

Table 1. Questionnaire developmental steps in three phases of a sequential mixed methods study

Phase	Step
Phase 1	Step one- Ethics and assembly of an expert panel
	Step two- Qualitative exploration of folk linguistic perspectives
	Step three- Components of the questionnaire and item generation
	Step four- Rating scales of the items
	Step five- Developing the preliminary demographic questions
Phase 2	Step six- Item checking via expert review
	Step seven- First pilot
	Step eight- Integration of data via methodological triangulation
	Step nine- Pilots 2 and 3
	Step ten- Validation
Phase 3	Step eleven- Administration of the validated questionnaire

1.2.2. Phase two: Quantitative strand- development of a measurement tool

In the second phase (see Table 1), based on the interviewees' linguistic perspectives, a review of available literature on language identification and the panel of experts' views, I developed and validated a contextualized questionnaire. As Creswell (2015) states, the impetus for the development of a contextualized tool is that measures drawn from other studies may have limited applicability in the context of the current study. The validation process of this questionnaire took nearly two years and eight steps to complete (see Table 2). Detailed descriptions of these steps are provided in part two of the dissertation.

1.2.3. Phase three: Quantitative strand- administration of the questionnaire

In the third phase, the language identification questionnaire helped address the issues of language identification in Iran more comprehensively and systematically. I administered, collected and analyzed 192 questionnaires to examine Iranian folk linguistic perspectives and further investigate whether their demographic information influenced their perspectives. As can be

noticed, Table 2 provides more information regarding the timeline in which I completed this research project.

Table 2. PhD project calendar- Exploratory mixed methods research project (2017-2021)

Year/Month Phase	2017		2018	2019	2020	2021							
	Jan.-Mar.	Jun.	Feb.-Dec.	May -Dec.	Jan.-Dec.	Jan.	Feb.	Mar.	Apr.	May	Jun.	July	Aug.
Phase one- Ethics approval and assembly of an expert panel (Step one)													
Phase one- Qualitative exploration of folk linguistic perspectives (Step two)													
Phase one- Presentation of qualitative results at ICKS3													
Phase two- Questionnaire development and three pilot iterations (Steps three-nine)													
Phase two- Questionnaire validation process (Step ten)													
Phase two- Paper one writing process													
Phase two- Paper one submitted to <i>LRR</i>													
Phase two- Paper one accepted with revision													
Phase three- Questionnaire administration (Step eleven)													
Phase three- Paper two writing process													
Phase three- Paper two submitted to <i>AILA</i> (under consideration)													
Hybrid PhD dissertation writing process													

1.3. Hybrid structure of this dissertation

The structure chosen for a dissertation depends fundamentally on the discipline and subject area of the research. Historically speaking, broad focuses and complex research questions are often addressed in the structure of a monograph PhD (Williams, Stevenson, et al., 2009). A monograph is a coherent specialized text, produced by a single author and written for a specialized academic audience (Thomson, 2002). This type of dissertation is structured in chapters and discusses a single topic thoroughly and exhaustively (Williams, Stevenson, et al., 2009). A critical characteristic of a monograph is its large size. This feature is what distinguishes this type of dissertation from research articles. Despite the popularity of a monograph dissertation in many universities, in academic disciplines “a dissertation book based on articles, i.e., a compilation thesis” (Fridlund, 2010, p. 144) has become an established standard. This type of dissertation, which has a similar structure to a monograph dissertation, contains a condensed theoretical framework followed by an empirical section which includes peer-reviewed articles (Fridlund, 2010). It is crucial that the articles present a unified research project. Fridlund (2010) further discusses one of the major differences between the two dissertation styles as supervision by professionals. While a monograph PhD requires a student to discuss the research with a small number of scholars in a supervisory committee, in a compilation PhD, the research process is scrutinized through the peer-review process of journals in addition to a supervisory committee.

Compliant with Carleton University’s policies in the Applied Linguistics and Discourse Studies (ALDS) program, and with the consent of and in collaboration with the supervisory committee, a PhD student can produce a compilation thesis also known as “the ‘integrated article,’

‘manuscript,’ ‘sandwich,’ or ‘chapter’ thesis” (ALDS, n.d.). Similar to many disciplines which establish this structure as an accepted approach to thesis production, in the ALDS program at Carleton University, the PhD students who pursue a manuscript PhD must author or co-author (as the lead researcher) three or more papers and present them in peer-reviewed venues. The ALDS program requires that the body of a manuscript thesis contain an abstract, preface, introductory chapter, methods chapter, data chapter (research papers), conclusion, and references.

The involvement of professional publishers in the process of writing a manuscript PhD, albeit with many advantages, entails challenges and limits posed on research in a worldwide pandemic. These challenges may bind a PhD student to work within a less desirable environment and delay the overall workflow of the dissertation. One such difficulty is lack of control exerted over the timeline during which a student wishes to complete the research project. As such, I developed an awareness of and accepted my limitations as a researcher in all the research-related steps.

The current dissertation, originally confirmed as a manuscript style, was challenged on the grounds of limited access to participants in completing the empirical section of this research project. This, in turn, prolonged and delayed the course of data collection and analysis, the writing process of research papers, and publication by nearly a year. What complicated matters more was the journals’ timelines in responding to submissions. Correspondence with some of the journals suited to submit language identification papers revealed that these timelines had been postponed by a few months.

Considering the circumstances, upon my supervisory team's approval, my research results are in a hybrid style dissertation. That is, this dissertation combines and benefits from the fundamental elements of a monograph dissertation as well as the features of a manuscript dissertation with academic works submitted for publication. This dissertation, which took nearly five years to complete, includes two main parts.

The first part, representative of the sections of a monograph thesis, provides a review of literature about different approaches taken in addressing the issues of language identification in Iran and other countries. Further, this literature review focuses on parameters (linguistic and non-linguistic) which are significantly associated with scholarly and folk perspectives in identifying language varieties. In addition to literature review sections, this first part contains conceptual framework and methodology of this dissertation. This part, in turn, underlines the significance of research carried out in the second part (empirical section) of this dissertation.

The second part contains the three phases of the exploratory sequential mixed methods research project. I completed these phases during multiple steps (see Table 1). The results obtained from the qualitative study was presented in a conference (ICKS3) and results of the two quantitative studies were respectively submitted to *Language Related Research (LRR)* (accepted with revisions) and *AILA* (under consideration) journals. Lastly, typical of a monograph dissertation, I present and discuss the conclusions and implications of this research project and close this dissertation with a list of references and appendices.

2. Hybrid thesis: Part one

This part begins with the literature review section. In this section (Section 2.1.), I elaborate on approaches taken to address the issues of language identification and underline the significance of an eclectic approach. This approach underscores that systematic identification of language varieties requires researchers to include linguistic perspectives of speech communities in addition to scholarly views. In the conceptual framework section (Section 2.2.), I draw on three theories in order to develop a measurement tool which can comprehensively address the issues of language identification. After this, in the methodology section (Section 2.3.), I describe the circumstances under which I recruited the participants and the techniques that I used to collect and analyze data.

2.1. Literature Review

In order to investigate, analyze and develop an understanding of folk perspectives about different types of language varieties, I discuss the scholarly approaches adopted to address the issues of language identification, specifically in the field of Iranian linguistics. Section 2.1.1 explores the approaches taken to address these issues in the past few decades. In this section, I start with studies which have focused on structural approaches to language identification and underlined the significance of linguistic features of language varieties. I continue with an elaboration on a functional approach which uses the measures of inherent mutual intelligibility in language identification. Lastly, I focus on the importance of ethnography of speaking which underlines the equal significance of non-linguistic parameters in language identification. This approach considers language as a social entity which cannot be examined detached from the context where it is used and people who use it. In response to the failure of much language attitude research to pose questions which address the language identification issues, Section 2.1.2 introduces the domain of folk linguistic enquiry. On a related topic, Section 2.1.3 identifies who the folk population is. Finally, Section 2.1.4 reviews the available literature on issues related to language studies in Iran.

2.1.1. Approaches to language identification

Studies of language have commonly analyzed spoken and written materials focusing on the structural features of language varieties. Chomsky, for example, in the main body of his earlier works (e.g., see Chomsky & Halle, 1968), detached language from actual speech and context of use. A Chomskyan approach to the study of language, which is reflected in the structural approaches taken by many Iranian linguists, does not adequately account for variation because it limits its focus to the potential language production of an idealized speaker/hearer as its object of study.

It is challenging to identify language varieties as being one particular type of entity or another, due to the ambiguities attached to terms like ‘language’ or ‘dialect’ (Saville-Troike, 2003). Taking a strictly structural perspective, some scholars – including many linguists in Iranian linguistics – postulate these terms as objectively distinguishable entities (Miller et al., 2006; Perry, 1998; Windfuhr, 2009). These linguistic terms might be confusing, but linguists have defended such distinctions as practical (e.g., see Bloomfield, 1933; Haugen, 1966). This approach to the study of language was particularly concerning to sociolinguists such as Hymes (1974). Hymes, in particular, criticized Chomsky’s failure to account for linguistic variation within speech communities.

In Iran, with the purpose of defining, differentiating and organizing language varieties, language identification studies (e.g., see Aliakbari et al., 2015; Borjian, 2008, 2013) have focused on the structural features of language varieties. This structural perspective equipped scholars with detailed descriptions of the lexical similarity, along with the phonological, morphological and

syntactic systems of the respective varieties. Furthermore, it successfully assisted scholars in producing classifications of language varieties (e.g., see Anonby, Taheri-Ardali, et al., 2015-2021; Windfuhr, 2006, 2009).

Rather than solely determining language as a set of rules, Hymes (1989) reiterates the significance of social context within which communication occurs. Linguistic investigation methods which resulted from Hymes's perspectives regarding linguistic variation is the ethnography of speaking (later amended to ethnography of communication) where shaping and analyzing speech is not dependent on linguistic accuracy (Hymes, 1974, 1989).

Although linguists often believe they can consistently define and differentiate a 'language' from a 'dialect', language varieties frequently merge into one another along parameters such as status of language varieties in the education system, government and social media (Modarresi, 2014; Moradi, 2019; Preston, 2016), geographical proximity (Chambers & Trudgill, 1998), community membership and sociocultural values (Fishman, 1965, 1972), religious and cultural identity (Joseph, 2004), linguistic labels (Dabir-Moghaddam, 2008; Modarresi, 2014), speech community members' age, gender, linguistic background, ethnicity, education, and profession (Saville-Troike, 2003), mutual intelligibility (Chambers & Trudgill, 1998; Eberhard et al., 2021; Gooskens, 2007), and even structural parameters such as the proportion of shared words, sounds, or grammatical structures (Modarresi, 2014; Wolfram, 2017). Thus, researchers need to realize that language varieties are not easily identifiable units with clear-cut borderlines (Makoni & Pennycook, 2007; May, 2012). As Eberhard et al. (2021) assert, "how one chooses to define a

language depends on the purposes one has in identifying one language as being distinct from another”.

In identifying different types of language varieties, Eberhard et al. (2021) apply three basic criteria. The first criterion identifies two language varieties as belonging to the same language based on a functional level of mutual intelligibility. This is to say, if the speakers of these language varieties can understand one another based on knowledge of their own variety and without learning the other variety, these two varieties are recognized as belonging to the same language. The second criterion underlines the significance of social parameters such as individuals’ ethnicity in their language use. When the speakers of two language varieties share a common ethnolinguistic identity, whether or not spoken intelligibility exists, these varieties are considered varieties of the same language. Conversely, the third criterion calls attention to the existence of distinct ethnolinguistic identities of the speakers of two language varieties in the face of sufficient mutual intelligibility. Under such circumstances, even when mutual intelligibility enables communication, these varieties should be considered different languages.

These criteria which are echoed in the writing of other scholars (e.g., see Mariou, 2017; May, 2012) have shown that despite mutual intelligibility among language varieties, due to social conflict, those varieties are sometimes referred to as different languages rather than varieties of the same language. Therefore, identification of language varieties can take place entirely independent of mutual intelligibility or structural similarity (Wolfram, 2017). Thus, adopting a functional approach where language varieties are identified based on mutual intelligibility may not be sufficient proof to determine if language varieties belong to one language rather than two separate

languages (Nettle, 1999; Saville-Troike, 2003). Nettle (1999) refers to mutual intelligibility as a parameter rarely used in linguistic-related studies, problematic in basis, and variable with respect to context and community members involved.

To elaborate on the inadequacy of this parameter alone in language identification, Nettle (1999) expands on the organization of language varieties into a dialect continuum. This type of continuum is a chain of language varieties spoken across geographical regions. As embedded in the definition, the linguistic differences between neighboring language varieties are cumulative. While the successive varieties are mutually intelligible, the further away from one language variety, the higher the chances of lesser intelligibility to the point that two varieties at opposite ends of the continuum are likely to be completely unintelligible. Therefore, the distance between language varieties from one another is the factor that can be closely correlated to mutual intelligibility (Chambers & Trudgill, 1998; Heeringa & Nerbonne, 2001; Nettle, 1999).

An example to demonstrate dialect continuum would be the Fula language spoken in nearly 20 different countries mostly in West and Central Africa. While speech community members from neighboring dialects understand each other mutually, a Fula speaker from one end of West Africa would probably find it problematic speaking with a Fula speaker from another dialect located towards the other end of the continuum. Despite geographic span, linguistic variation and lack of mutual intelligibility between varieties of Fula, some commentators consider them to be dialects of the same language (Ka, 1991). Nonetheless, despite lack of sharp boundaries, Eberhard et al. (2021) treat these language varieties as separate languages. In this language situation, identifying

bits of a dialect continuum from a purely linguistic or functional point of view must be at least partially arbitrary (Heeringa & Nerbonne, 2001).

Considering the inadequacies of a functional approach to language identification, scholars underline social identification rather than mutual intelligibility as a more significant parameter which decides whether individuals want to be identified together (LePage, 1968; Saviile-Troike, 2003; Wolfram, 2017). In such cases, whether mutually intelligible or distinct, speech community members adjust their linguistic behavior, which would result in downplaying the unique features of those language varieties.

Ethnographers identify language as an inseparable part of the interaction process where it communicates more than what the surface words seem to imply. Gumperz's (1970) and Hymes' (1967) ethnographic approaches to the study of language deem the total context of communication just as worthy of study as linguistic performance data. An ethnographic approach to the study of language focuses on meaning conveyed with more than single sentences which conform to the rules of grammar. This approach underlines that meaning is conveyed in complex exchanges where folk knowledge of a particular social situation and community determines how they use language. As such, an ethnographic approach can develop a better understanding of the underlying reasons why non-linguists, possessing agency as members of specific speech communities, hold the perceptions that they do (e.g., see Niedzielski, 1999; Sabethematabadi, et al., 2017).

In addition to identifying parameters which influence definition and differentiation of different types of language varieties (Lavandera, 1988; Saviile-Troike, 2003), linguists have increasingly realized that there are differences of perspectives between speech community

members' and scholars' approaches to language identification (Anonby & Sabethemmatbadi, 2019; Niedzielski and Preston, 2000). These perspectives underline the important role of people who speak various language varieties not just as collaborators of research where they are treated as informants or subjects of empirical studies, but also as sources of linguistic knowledge (Leonard & Haynes, 2010; Mithun, 2007).

Identification of language varieties “requires observation and description of actual differences in pronunciation, grammar, lexicon [...], but it must ultimately depend on the discovery of which differences are recognized by members of the group as conveying social meaning of some kind” (Saville-Troike, 2003, p. 41). Therefore, scholars should be mindful and inclusive of individuals' linguistic perspectives in addressing the issues of language identification. Cook (2003) considers differences of perspectives between the experts and non-experts as not easily aligned and even likely “to be aggravated by an attempt to impose insensitively an ‘expert’ view which runs contrary to deeply held belief” (p. 12). For richer interpretations of properties of languages and to compensate for the imbalance between researchers' and speech communities' contributions, not to mention the validation of cultural beliefs and practices, numerous publications have encouraged a movement away from a research-centric model with a traditional perspective of speech communities as collaborators in data collection (e.g., see Amery, 2009; Grinevald & Sinha, 2016; Hill, 2002).

From the speech community members' perspectives, defining language varieties depends greatly on how speakers are able to preserve their heritage and culture, and maintain the unity and coherence of their community using language as a means to this end (Saville-Troike, 2003;

Eberhard et al., 2021). The number of parameters, and the priority given to them, differs from one group to the next inside a speech community, not to mention from one speech community to the next. Hudson (2001) argues that due to such complexities, not only are there great challenges in delimiting one variety from the other variety of the same kind (e.g., one dialect from another dialect), but also in delimiting one kind of variety from the other (e.g., language and dialect).

2.1.2. The significance of folk research

As Preston (1999a) notes, identification of language varieties from people's perspectives and ways in which their perceptions are arrived at – the subject area of folk linguistics – are of fundamental importance, yet for the most part poorly examined even today. Scholars have persistently discounted folk beliefs about language as invalid and, in some cases, designated these as 'secondary and tertiary responses' (see Bloomfield, 1944). These scholars have considered folk perspectives about language varieties as "at best, innocent misunderstandings of language, at worst, the bases of prejudice, leading to the continuation, reformulation, rationalization, justification, and even the development of a variety of social justices" (Niedzielski & Preston, 2000, p. 2). As Preston (1989) states, this could be partly due to Labov's (1972, p. 113) notion of 'observers' paradox' which suggests that the more aware participants are that their linguistic behaviour is being observed the less likely it is that the behaviour will be natural.

As Preston (1989) explains, "it has been hard to imagine naïve metalinguistic concepts of the ordinary, well-educated non-linguists among us constitute valid social scientific information" (p. xi). Folk linguistic data has been dismissed as 'impoverished', since non-specialists are thought to have an inadequate vocabulary with which to discuss their thoughts on language variation

(Preston, 1989). Niedzielski and Preston (2000) further reject the idea of the inadequacy of people's remarks on their own language, this so-called 'impoverishment of data'. They underline that the focus of language studies is often limited to linguistic features, especially phonological features of the language varieties under study. In such studies, scholars such as Labov (1972) have not paid attention to people's attitudes, notions of language change, acquisition, multilingualism, and intelligibility. Nor do they appeal to sociolinguistic evidence such as region, age, ethnicity, or status in their rejection of people's beliefs. Abandoning the people's knowledge because of limitations in their technical knowledge of linguistics seems unwarranted.

Goossens (1970, as cited in Preston, 1999a) argues that due to sociopsychological realities associated with people's perceptions of language varieties, their judgements may not always correspond to those of linguists and this difference may, in turn, impede the progress of language studies in general. Furthermore, it may be challenging to control and measure sociopsychological elements that enter into their judgements (Preston, 1999a). However, no matter how challenging the measurement and collection of folk data may be, without knowledge of how their perceptions address linguistic realities, language studies are no more than "a venture into the investigation of academic distinctions which distort the folk reality or tell only a partial truth or, worse, a misadventure into the study of theatrically exaggerated speech caricatures" (Preston, 1993, p. 252).

Preston (1989) argues that studies which probe folk perceptions and beliefs about language varieties cannot succeed merely at a micro level, fixated on linguistic properties of language varieties. The macro-linguistic (ethnographic) perception studies which seek people's understanding of language variation can help reflect a holistic image of the varieties spoken in

particular geographical locations. He adds that such studies, should they be appropriately conducted considering the observer's paradox (Labov, 1972), can investigate speakers' taxonomy of language variation categories, social characteristics provided by speakers as proof for linguistic differences, the geographical areas where they believe the linguistic differences exist, and perceived values of language varieties. In analyzing a folk linguistic perspective, the empirical and scientific accuracy of folk linguistic perspectives is not of first importance. What matters is that the analysis of non-specialists' views reflects their own world and shows patterns and consistencies representative of the wider world (Albury, 2017; Preston, 1994).

With respect to the prevalence of folk knowledge, various scholars (e.g., see Benson, 2003; Eppler & Benedikt, 2017; Preston, 1999c; Wilton & Stegu, 2011) note that when asked by linguists, people share their opinions about language learning, language peculiarities of other people, communication problems, and language use in the media. Preston (2005) categorizes this type of folk knowledge as 'suggestible knowledge' which is not readily available as the topic of everyday conversations, yet folk access it when they are required to.

Folk views about language varieties and their distribution are profoundly beneficial to the practitioners who work in unfamiliar speech communities (Preston, 2011). Furthermore, attempting to understand degrees of similarity between different language varieties spoken in various geographical areas without access to the perspectives of people who speak the respective varieties may impede the efforts of an outsider linguist, a language planner or a policy maker. From a scientific view, taking folk truth value seriously in identifying language varieties does not prevent the researcher from consulting other experts on language. It simply means that

investigating their views is as essential a complement to scholarly study of language and language use (Wilton & Stegu, 2011). For example, such an approach to studying language yields specific, useful insights into why language diversity is declining at such a rapid pace (Anonby et al., 2019).

Folk dialectological studies have been most prominently conducted in the UK (e.g., see Inoue, 1996; Coupland et al., 1999), France (e.g., see Kuiper, 1999), Germany (e.g., see Kremer, 1984/1999), the Netherlands (e.g., see Rensink, 1955/1999), Japan (e.g., see Grootaers, 1964/1999; Sibata, 1959/1999), and the US (e.g., see Preston, 1999a, 2011). Such studies consider beliefs, views, attitudes, and theories of everyday language users as valid objects of academic investigation.

To determine dialect boundaries, Rensink (1955/1999) explored the degree of differences that the participants felt existed between their areas of residence and nearby areas with respect to dialects spoken in the respective areas. Based on the participants' answers, Rensink created a classification map of dialects which she later regarded as a tentative sketch which was only partially successful due to missing and contradictory data as well as lack of clear boundaries.

To investigate folk dialect consciousness or awareness, Weijnen (1946/1999) devised the 'little arrow method' which connected the participants' home areas to geographically nearby areas that they underlined as linguistically similar. Such maps provide opportunities for folk linguists to examine the degrees of similarities between traditional dialect divisions and those perceived by the participants. Though no perfect match exists in these maps, the perceptual-production matches are impressive. In a similar work in Germany, to explore whether the Netherlands-German national border appeared as a linguistic border in the minds of speakers, Kremer (1984/1999) investigated

subjective dialect boundaries using the same method introduced by Weijnen. In that study, Kremer asked the participants whether dialects spoken in areas near them were similar to theirs. In cases where similarities existed, dialects were connected by means of an arrow (either one-headed or double-headed) on a map creating a net or chain. The results revealed that the Netherlands part “contained more one-headed arrows, an occurrence which points to a greater uncertainty of the informants or to smaller differences between the dialects” (p. 35). In Japan, Nomoto (1963/1999) investigated whether the informants perceived any of the twelve survey areas located in a straight line as being the same as their own speech. Results revealed that informants’ perception of linguistic similarities between their own and others’ speech in the respective survey locations corresponded to natural (mountains and rivers) and political boundaries.

Whereas some folk dialectologists (e.g., see Kremer, 1984/1999; Rensink, 1955/1999; Weijnen, 1946/1999) have required people to draw a line from their own language variety to those which are linguistically most similar (the “arrow method”), others (e.g., see Grootaers, 1964/1999; Sibata, 1959/1999) have determined the subjective boundaries of dialects and the degree to which these varieties were perceived to be different from one another. Partially inspired by the work in the Netherlands, Sibata (1959/1999) requested his informants to identify whether nearby villages were ‘not different’, ‘a little different’, ‘quite different’, or ‘incomprehensible’. The resulting subjective boundaries did not correspond to the traditional dialect boundaries in those areas. As such, Sibata claimed, and later Grootaers (1964/1999) confirmed, that such results may be of little interest to linguists. Nonetheless, Grootaers was more inclined to attach significance to non-linguistic parameters in dialect boundaries. Hammarström (1961, as cited in Grootaers, 1964/1999)

established that it is only legitimate if instead of the subjective choices of the researchers who determine the objective boundaries, one chooses those boundaries set by the very people who use the language. This is to say that in folk dialectological applications, folk linguists advocate an approach to studying speech which establishes “the traits of language upon which subjective dialect differences rest” (Grootaers, 1964/1999, p. 119).

Only after linguists are able to access folk perspectives about dialect differences, regional speech zones, and the characterizations of language varieties in general, are they able to conduct studies which are representative of public discourse as well as academic facts (Preston, 1999a). Building on the studies carried out between the 1950s and 1970s, a group of folk studies investigated folk perceptions of language varieties and principal characteristics associated with these varieties (e.g., see Ryan et al., 1982). These studies which determined folk perceptions of regional varieties, drew on techniques such as hand-drawn perceptual maps (Inoue, 1996), pleasant and correct assessments (Kuiper, 1999), and degrees of difference ratings (Sibata, 1959/1999), all of which underlined folk demographics such as age and gender as well as the status of language varieties as significant in their findings.

As an example, in his study, Inoue (1996) studied the socio-psychological images of various dialects and accents in Japan by asking his informants to apply evaluative words to their own varieties. He believed that language varieties should be treated as social phenomena with emotional aspects worthy of investigation. His findings confirmed dialect image as a plausible technique and dialect inferiority complex, that is, discriminatory treatments of language varieties by people and governments, as a real issue which requires further attention. In a similar study,

Long (1999) examined the perceptual dialect maps drawn by ordinary people in Japan. Results revealed that the informants' maps did not always match up with cultural and political divisions. That said, Long believed that more often than not, folk dialect maps coincide with cultural boundaries.

Using a different method, namely: "pleasant and correct" assessments, Dailey-O'Cain (1997/1999) examined informants' attitude towards the correctness and pleasantness of language varieties of 36 German regions. Results indicated that the Western informants found the Western varieties more correct and pleasant than the Eastern varieties. That said, the Eastern informants did not perceive any difference in terms of correctness while agreed with the Westerners about the pleasantness of Western over Eastern varieties. She concluded that these subjective assessments influenced other aspects of the informants' lives such as educational and professional opportunities. Thus, it is important that people develop awareness regarding bias arising from these differences in language attitudes which may be driven by social and political motivations.

The modern trend in perceptual dialectology continued with Preston's (1999b, 1999c) own work via multiple techniques. These techniques including draw-a-map, degree of difference, correct and pleasant, and dialect identification have been used in the United States (e.g., see Benson, 2003; Hartley, 1999; Preston, 1989) as well as around the globe (see this section above).

Draw-a-map tasks (see Preston, 1982; Montgomery, 2007) collect the respondents' impression of the environment that they and/or other speakers live in by adding lines onto a blank map where they believe dialect boundaries exist. Adopting this technique, Montgomery (2007) conducted multiple pilot studies. In each of these pilots, a different panel of respondents was given

maps with the outlines of England and Wales. In the first pilot, the panel of respondents was asked to draw lines around dialect areas. In the following pilots, however, not only were the respondents given the respective maps, but they also had access to a separate map with the location of the major cities. The results indicated that the respondents' recognition rate increased from the first to the second and third pilot iterations with access to the second map. Though a second map increased the confidence of the respondents, the addition of the location map with major cities did not influence the number of dialect areas added onto the blank maps. Similar studies (e.g., see Benson, 2003; Cramer, 2010) used this technique and provided the respondents with more or less the same basic outline maps. Common among these studies were questions concerning the folk perception of where dialect areas exist and how they perceive the extent to which varieties spoken in other areas are similar to or different from theirs.

Included among perceptual dialectological studies are those which collect data regarding folk beliefs about language in general (e.g., see Inoue, 1996/1999; Preston, 2013; Williams et al., 1999). Such studies capture language variation in terms of their social features such as pleasantness and correctness. Further, these rich databases help dialectologists and linguistic geographers develop a better understanding of language varieties that folk consider as more desirable, pleasant, useful, or correct compared with others.

Though scholars admit that it has been only within the past few decades that studies have focused on folk attitudes and beliefs about language varieties in certain areas and with respect to limited techniques (Hartley, 1999), such academic movements are overdue in Iran. Among possible concerns, valid as they may be, is the accuracy of data with respect to these beliefs.

However, as Preston (2005) asserts, being accurate or not has no bearing on the value of this sort of data. Providing a comprehensive picture of the language situation in any given country is not possible until folk data has been collected from at least major geographical areas of the country (Hartley, 1999).

In Iran, scholars have merely provided scholastic definitions for linguistic terms and reached a consensus regarding these terms, albeit minimal, to determine the borderlines between language varieties. Based on these agreements (see Section 2.1.4.), they have generated classifications and documentations of language varieties spoken in Iran. Whether a variety is considered as belonging to one category or another is inextricably tied to linguists' perspectives and their approaches to studying language. Rarely did they investigate whether the subjective accounts of members of language communities align with those of linguists or the underlying reasons for such alignments, if any.

2.1.3. Who folk are

One of the problems that folk linguists focus on is a proper definition for who folk are. As valuable as the addition of folk perspectives to the field of linguistics is, investigating these perspectives regarding patterns of language identification gets complicated as the groups whose views are informed by either academic or folk knowledge are not clearly distinguishable. As folk linguists (e.g., see Niedzielski & Preston, 2000; Wilton & Stegu, 2011) assert, the dichotomy of linguists/non-linguists is not presented with clear-cut boundaries, and rather develops along a continuum. The reason is that the knowledge which informs both groups is not academic versus folk. In the same way as the scholars' source of knowledge is academically and empirically

substantiated information, folk may have access to various other resources which inform their views. As such, folk linguists typically advocate a linguistic perspective which is more hybrid in nature, encompassing both academic and folk views.

Paveau (2011) suggests that being folk is an activity rather than a permanent status due to unclear boundaries and differences among those included in the category of folk. Although most people with different types of profession probably do not have technical linguistic knowledge (Niedzielski & Preston, 2000), data collected from each of these language users may vary. For example, the value of information which lawyers provide is different from that of a person who is not literate (Paveau, 2011).

Therefore, in defining who folk are, the idea of a continuum with permeable and negotiable boundaries is more appropriate when it comes to language issues. In such a continuum, at one end, are people with no linguistic knowledge or training, and at the other end there are people who, due to interest in linguistics and/or via media and education have received some formal knowledge of linguistics. On a similar note, Wilton & Stegu (2011) assert that ‘folk’ cannot be defined separately from their social environment and their roles. Thus, folk perspectives about language varieties cannot be viewed as merely folk or academic, but rather a hybrid of both with regard to their discourse.

Preston (2011), on the other hand, suggests that except academic linguists everyone is in the category of folk. He also adds that he does not use the term to refer to “rural, marginalized, less educated, or romanticized (‘quaint’) groups. We’re all folk when we step into the world of traditional knowledge and ways of behaving outside our own technical training” (p. 15). These

inexpert linguists need not rely on any qualification in linguistics as discussion and application of linguistic knowledge are not exclusively the activities of trained linguists.

Folk perspectives about and descriptions of language varieties may not bear much resemblance to scholars' perceptions of linguistic reality. This is to say that the approaches that folk adopt to identify language varieties may not always be in terms of linguistic labels used by specialists, but rather descriptions of how they perceive the varieties (Preston, 2011). Folk descriptions define and classify language varieties in terms of their pleasantness, formality and usefulness (Preston, 1999a, 2011). This could be, among other factors, due to extralinguistic priorities such as social status, sociocultural norms, educational background, and so on (Preston, 2011; Saville-Troike, 2003; Wardhaugh, 2006).

Cramer (2016a) argues that unlike what some linguists (see e.g., Bloomfield, 1944; Labov, 1972) believe about folk (non-linguists), folk themselves are aware of the variation in their language and capable of discussing it. This provides an opportunity for linguists to examine the non-linguistic patterns beyond folk perceptions, since “the inclusion of folk beliefs in linguistic studies can offer important insights into the realities of language variation and change” (Cramer, 2016a, p. 2).

Considering the lack of consensus among folk linguists with respect to folk definition, in this study, I adopted the idea of a folk continuum (Niedzielski & Preston, 2000) which includes all non-linguists whose perspectives are informed by various sources, academic or otherwise. As folk views may be influenced differently given the impact of the unique sociolinguistic

associations of their speech communities, it is challenging to establish clear boundaries along this continuum.

2.1.4. Aspects related to language studies in Iran

Iran is home to many ethnolinguistic groups and languages. Thus, it is one of the most interesting linguistic landscapes in the Middle East (Moradi, 2019). Although official administrative records indicate between four and seven languages (Anonby et al., 2016), other suggest that approximately 80 language varieties are spoken in Iran (Rezaei & Bahrami, 2019). As the official language of the country, Farsi (also commonly referred to as Persian in English) is spoken by nearly 62% of the population in Iran (Izady, 2014). Next to Farsi, several other language varieties including Azeri, Kurdish, Gilaki, Mazandarani, Lori, Arabic, Balochi, and many other language varieties are spoken in Iran (Moradi, 2019). Due to migration, intermarriage and the priority of Farsi in the education system and media in Iran, many of these minority speakers either adopt Farsi as their second language or switch to it altogether (Izady, 2014). As such, the population of Azeri and Kurdish speakers as the two largest minority speech communities has decreased by nearly 2% between 1964 and 2014 (Izady, 2014).

As the largest ethnic minority, nearly 20 million people of Iran's population of 80 million speak Azeri as their mother tongue (Rezaei et al., 2017). Major dialects of this language include Tabrizi, Urumia'i, Ardebili, and Zanjani (Windfuhr, 2012). Due to mass migration, it is challenging to determine all of the specific Iranian territories where the Azeri minority speakers live. However, the majority of these speakers live in northwestern Iran (Moradi, 2019; Rezaei et al., 2017).

As the second-largest minority language community, Kurds make up nearly 11%-15% of Iran's population (Sheyholislami & Sharifi, 2016). Although some scholars (e.g., see Hassanpour, 1992; Sheyholislami, 2012, 2019) refer to Kurdish as a single language, McCarus (2009) identifies Kurdish as “a cover term for the largest group of closely-related Western Iranian dialects, spoken in a large contiguous area that extends from Turkey into Iraq, Iran, and Armenia, with smaller groups in other countries” (p. 587). In Iran, most of the Kurdish speakers live in Ilam, Kermanshah, Kordestan, West Azerbaijan, and Khorasan provinces (Sheyholislami, 2012). The main Kurdish dialect groups are listed as Northern Kurdish (Kurmanji), Central Kurdish (Sorani), Southern Kurdish (Kermanshahi), Hawrami (Gorani), and Zazaki (Dimilki) (Hassanpour, 2012; Rezaei & Bahrami, 2019). Although all of these dialect groups except for Zazaki are spoken in Iran, the majority of Kurds in Iran speak Central Kurdish. It is worth noting that while Kurdish does not have an official status in Iran or Turkey, it is an official language in Iraq and the main language of public administration, education and media in Iraqi Kurdistan (Sheyholislami, 2015).

Gilaki and Mazandarani speakers constitute approximately 10% of the population in Iran and live in Gilan and Mazandaran provinces, respectively (Moradi, 2019). As varieties spoken mostly in rural areas of the country, due to lack of sociopolitical movements which could maintain their survival, Gilaki and Mazandarani have not enjoyed the same level of vitality as the other minority languages in Iran such as Kurdish or Azeri (Moradi, 2019). Stilo (1981) refers to Gilaki and Mazandarani as two closely related languages with similar vocabularies, which share certain features with Caucasian languages, indicative of their close relatedness to the Caucasus area. In

the Atlas of the Languages of Iran (ALI), however, these two language varieties are classified as belonging to the Caspian language family, one of the six branches under West Iranian languages.

Balochi is another of the larger unofficial minority language families (Jahani & Korn, 2009) which is spoken by nearly 2% of the population in Iran (Tohidi, 2009). In the classification of Balochi, Jahani and Korn (2009) provide four major language varieties including Southern, Western and Eastern Balochi as well as Koroshi. Jahani (2001) asserts that Balochi is spoken in the southeastern corner of Iran in Sistan and Balochistan province. However, historically speaking, it is a Northwestern Iranian language (Jahani & Korn, 2009). They explain that the northwestern part of Iran is Balochi speakers' original habitat. These speakers immigrated to the southeast "under pressure from the Arabic and Turkic invasions of the Iranian plateau" (Jahani & Korn, 2009, p. 634). They further argue that all the established varieties of Balochi are geolects rather than sociolects or varieties based on factors other than geographical distribution. Therefore, until systematic studies are carried out, various Balochi registers remain unknown.

While these speech communities are the ones featured in this study, many other languages are spoken across Iran: to name a few, the Iranian languages Tati and Raji, the Turkic languages Turkmen and Khalaj, Neo-Aramaic, Armenian, Georgian, the Dravidian language Brahui, and Persian Sign Language (Anonby, Taheri-Ardali, et al., 2015-2021; Izady, 2014). Such language diversity and complexity in Iran has inspired many scholars to develop language atlases, maps and classifications over the past few decades. Despite the lack of consensus on dialect groups, over the years, scholars have created various classifications regarding geographical, historical and linguistic developments.

Since the 1950s, as part of a larger effort to document language varieties spoken in Iran, scholars such as contributors to Schmitt (1989), Stilo (2007b, 2016), Windfuhr (2009), Anonby, Taheri-Ardali, et al. (2015-2021) have developed classifications. Among the available classifications, most scholars pay attention to a comparative-historical classification of minority languages such as Balochi in Iran (see Jahani & Korn, 2009). Other classifications such as *the Atlas of the Languages of Iran (ALI)* highlight a complementary range of ways in which language may be classified: “genealogical inheritance, structural similarity through contact, and ethnic identification” (2015-2021).

In his linguistic composition of Iran, partly similar to *the Atlas of the languages of Iran (ALI)*, Izady (2014) divides the language families spoken in Iran into Altaic family (Turkic branch), Semitic family, Dravidian family, Iranian branch of Indo-European family, and Armenian branch of Indo-European family. In their ongoing work, Anonby, Taheri-Ardali, et al. (2015-2021) group language varieties according to eight different language families to generate a classification of all the varieties spoken in Iran. This atlas, *the Atlas of the Languages of Iran (ALI)*, includes Iranian, Indic, Armenian, Turkic, Semitic, Kartvelian, Dravidian, and sign language. Further, it has congregated an inventory of 500 language varieties spoken in Iran with almost 400 of them belonging to Iranian language family. The largest and most detailed branch of New Iranian period is what Anonby, Taheri-Ardali, et al. (2015-2021) label ‘West Iranian’. This language family consists of six primary sub-groups: Kurdish, Caspian, Gorgāni, Northwestern, Balochi, and Southwestern.

Language mapping also accomplishes the function of social representation, inventory and organization, albeit in a less explicit way. One important attempt to map the languages of Iran was

the atlas project initiated by Georges Redard during the 1950s and 1960s but discontinued after ten years (Redard, 1974). The Persian Academy also carried out work during the 1970s. This project was restarted after 2001, but the release of results in the form of a language atlas is still awaited (the history and contributions of this project are discussed in: Akhlaghi, 2020; and Taheri-Ardali et al., 2021). Additionally, map series such as Atlas Narodov Mira (Buk & Apenchenko, 1964), TAVO (Orywal, 1988), Izady (2014), and Irancarto (Hourcade et al., 2012) which expressed dominant ideas about language taxonomies in Iran, were based on very general classifications and faced significant issues of scope and consistency.

A review of available maps, atlases and classifications of Iranian language varieties reveals that rarely do classificatory and documentary projects work systematically or address adequately the complicated linguistic and social realities in Iran from the perspective of its people, which would provide an overview of language use, language distribution and identity, and would accommodate linguistic ideologies of speech communities next to those of scholars and policy makers. Apart from initial research on perceptual dialectology of Kurdish-speaking areas extending into Iran (Eppler & Benedikt, 2017), folk linguistics has been neglected as an area of research in Iran. Hassanpour (1998), however, considers the perspectives of speech community members as possessing the same level of authority as those of linguists.

As stated earlier, language identification research in Iran reveals that linguists are identifying language varieties and types of language varieties according to their own perspectives and criteria. What complicates matters even more is that parameters associated with language varieties, whether linguistic or non-linguistic, are neither compatible nor constant across speech

communities (Chambers & Trudgill, 1998; Modarresi, 2014). For example, two language varieties may be considered two separate languages based on the extent of mutual intelligibility, the proportion of shared vocabulary, or the existence of major sound changes that show that they have evolved independently. However, the same varieties may instead be perceived as varieties of a single language on the basis of cultural or political factors. Under such circumstances, definition and differentiation of different types of language varieties are quite challenging (Modarresi, 2014).

Although some Iranian scholars primarily focus on linguistic features and mutual intelligibility as the two main parameters in identifying language varieties, scholars of Kurdish origin (e.g., see Hassanpour, 1992, 1998; Sheyholislami, 2015, 2019) underline the significant contribution of sociocultural and political parameters. Such sociopolitical forces are responsible for speech community members' bilingualism but also, in some cases, their willingness to abandon their variety in favor of the official language, Farsi. As scholars (e.g., see Bani-Shoraka, 2005; Hassanpour 1992; Jahani, 2005; Sheyholislami 2012) assert, prior to 1979, not only politicians but also even many academics considered almost all of the languages of Iran, even those with very different genetic roots from Farsi, to be dialects of Farsi. These sociopolitical pressures towards a monolithic state language might undermine and even erase the lifestyle, culture and language of a minority group (Jahani, 2005) and lead to “the suppression of academic study of the language, its dialects, geography, and history” (Hassanpour, 1991, p. 33). Yet the condition of minority languages is not straight-forward; the responses to societal pressures towards a monolingual state language are complex and more nuanced than simple acceptance or rejection of a less favoured language, as the following examples will attest.

In recognizing the impact of these parameters on identification, use and status of minority languages such as Kurdish, scholars (e.g., see Sheyholislami, 2012, 2015; Sheyholislami & Sharifi, 2016) explain that despite its ranking as the third ethnic language community in Iran in terms of the number of speakers, Kurdish has neither been recognized as a language with an official status nor has it been exempt from suppression and endangerment.

Further, attempts to offer and implement studies about or in minority languages such as Kurdish have not been considered by the government except in rare cases such as the establishment of a research institute at the University of Kurdistan (Rezaei & Bahrami, 2019). Even though article 15 of the Constitution of Islamic Republic of Iran allows the use of language varieties other than Farsi at educational settings, as Sheyholislami (2012) explains, the Kurds' efforts for the use of their language as the medium of education at schools and communication in government offices in the Kurdish areas remain unproductive. Mirshahidi (2017) explains that the scarcity of research on topics related to minority languages in post-revolution Iran is because such studies are subject to government review. Considering these challenges in identification and use of minority languages (e.g., Kurdish) in the education and in government (see Sheyholislami, 2012, 2015, 2019), and in pursuit of better educational and vocational opportunities, it is not surprising that many minority language families are more inclined to speak in Farsi with their children at home.

In identifying Kurdish, Sheyholislami (2015) refers to further prominent challenges such as “lack of a common standard language, a unified writing system and mutual intelligibility across the main dialect groups” (p. 30). Although Kurdish has been recognized as a language by many linguists, they have not reached a consensus regarding its dialect groups (Haig & Öpengin, 2014;

Hassanpour, 2012; Skutnabb-Kangas et al., 2012; Shahiditabar & Pourghasemian, 2016). However, none of the diverse classifications of Kurdish dialect groups can disregard the fact that no matter how distinct linguists consider these varieties, the speech community members who speak these dialects identify themselves as Kurds and their language as Kurdish (Hassanpour, 1998).

Iranian studies conducted in recent years, albeit limited in number, have confirmed the vitality of minority languages in Iran and the continuing positive attitude of native speakers towards their mother tongue. Common among these studies is people's motivation to speak and maintain their heritage language as a part of their ethnic identity and connection to their ancestors. For example, Mirvahedi (2017) who investigates family language policies among Azerbaijani-speaking families describes how parents do not have a negative attitude towards the use of their mother tongue and use both Farsi and Azeri at home. In a similar study, in their investigation of the attitude of the Armenian native speakers in Isfahan, Rezaei and Farnia (2016) point out that Armenian speakers consider their language variety as part of their identity and use it in daily conversations. In addition, Rezaei and Tadayyon (2018) underline Armenian as one of the three languages used in the communications of Armenians in Jolfa (Julfa) besides Farsi and English.

The Iranian government has, to very limited extents, tolerated and allowed studies in and about the minority languages spoken in Iran. Establishing a research institute at the University of Kurdistan in Sanandaj to undertake Kurdish studies is one such step. Although programs of this nature help with the maintenance of minority languages in Iran, minority speakers can benefit from the merits of both their mother tongue and Farsi in a well-rounded education. After all, it is

undeniable that Farsi is still the dominant language of education, government, social media, and even the daily communication of many minority speakers (e.g., see (Rezaei & Bahrami, 2019).

Next to sociopolitical influences which complicate the identification of different types of language varieties, linguists themselves take different approaches in defining and differentiating the linguistic categories of ‘language’, ‘dialect’ and ‘accent’ (Dabir-Moghaddam, 2008). Such differing and, at times, opposing approaches have caused confusion and concern for Iranian linguists. According to Dabir-Moghaddam (2008), in the Iranian context much of this confusion is due to competing terms for ‘dialect’ as an intermediate category. ‘Dialect’ has been defined and interpreted differently in various Iranian publications (Batani, 2014; Dabir-Moghaddam, 2008; Modarresi, 2014). Dabir-Moghaddam (2008) points out that the Farsi equivalent of linguistic terms ‘dialect’ (Farsi=F. *guyesh*) and ‘dialectology’ (F. *guyeshshenāsi*) were coined by Kia (1961). Prior to Kia’s proposed equivalents, scholars and linguists in the field of Iranian linguistics expressed the idea of ‘dialect’ with a previously existing term *lahjeh*.

Many of the Iranian scholars identify different types of language varieties with respect to mutual intelligibility. In an attempt to define *guyesh* (which Farsi speakers tend to translate into English as ‘dialect’) and to differentiate it from *lahjeh* (more commonly translated as ‘accent’ today), scholars (Modarresi, 2014; Sadeghi, 1970) propose that *guyesh* is a language variety which is not easily understood by the speakers of Tehrani Farsi, a colloquial norm spoken in Tehran. Historically speaking, many of these varieties belong to the Iranian (Iranic) language family, but due to linguistic differentiation caused by geographical and temporal separation over millennia, they have diverged in fundamental ways. At the same time, speakers of these varieties – and the

linguists who share their perspectives – retain an understanding of some relationship to Farsi, and so these varieties are considered *guyesh*, and are not viewed on an equal footing with Tehrani Farsi by speech communities. Coupland (1988) underlines dialects as varieties not unambiguously associated with several social and regional variations simultaneously. Wolfram (2017) interpreted such flexibility as problematic in dialect definition.

Batani (2014) is one of the Iranian scholars who defines and differentiates different types of language varieties in terms of mutual intelligibility. Again, taking Farsi as a point of reference, Batani defines *lahjeh* as a variety of Farsi which is easy for speakers of other language varieties of the same *zabān* (roughly equivalent to the English word for ‘language’) to understand. However, if speakers of two language varieties do not understand each other at all, these language varieties must be considered two separate *zabān* (Batani, 2014). While Batani defines *lahjeh* and *zabān* with respect to mutual intelligibility, he identifies *guyesh* in terms of a writing system and its sociopolitical status. Regarding this, he underlines *guyesh* as an unofficial language spoken inside the political borders of a country. He further adds that, generally speaking, *guyesh* is a variety that has neither a separate literature or writing system, nor the same sociopolitical status as the official language of the country. Batani² refers to Kurdish, Gilaki, Mazandarani, and Balochi as such varieties.

Modarresi (2014) defines *guyesh* as a variety which differs in terms of grammar, vocabulary and pronunciation from other varieties whereas *lahjeh* is a language variety which is

² Please note that Batani’s (2014) book is a collection of old papers. He was not of the same opinion in later years.

only different from other varieties in terms of pronunciation. Like Bateni (2014), he calls attention to the significance of parameters beyond linguistic structure in defining and differentiating different types of language varieties.

Modarresi (2014) identifies mutual intelligibility as the most prominent and common non-structural parameter in defining language varieties. Mutual intelligibility between speakers of different language varieties can be helpful in identifying a borderline among the respective varieties. Nonetheless, considering that language is a sociopolitical and cultural phenomenon (Hymes, 1964, 1967; Saville-Troike, 2003), non-linguistic parameters may impact the extent of mutual intelligibility between the speakers of two language varieties. As such, there is no single parameter that can determine the borderline among language varieties.

Dabir-Moghaddam's (2008) views in defining different types of language varieties are similar to Modarresi's (2014), but are even more firmly grounded in the idea of mutual intelligibility. He maintains that the extent of mutual intelligibility between speakers of two *zabān* is zero. The speakers of two *guyesh* which belong to the same *zabān* mutually understand each other despite phonological, vocabulary and/or grammatical differences between their language varieties. The term *lahjeh*, again, is relevant for language varieties where speakers mutually understand each other with only phonological differences to distinguish the varieties.

In addition to differing perspectives among Iranian linguists, a heterogeneous populace of various religious and ethnic backgrounds complicates how the issues of language identification are addressed. A slight majority of people in Iran have mother tongues which they and others refer to as Farsi, but from a structural, historical and intelligibility-based perspective, these are often

distinct languages which belong to the Iranian (Iranic) language family and are unified through a shared cultural identity as Farsi (see Windfuhr, 2009).

The diversity and intricacy of language identification issues are augmented by the fact that not only do Iranians speak different Iranian (Iranic) language varieties, but a significant proportion of the population speaks language varieties which belong to other families: Turkic, Semitic, Indic, Armenian, Kartvelian, Dravidian, and Persian Sign Language (Anonby, Hayes, et al., 2020; Anonby, Sabethematabadi, et al., 2016; Orywal, 1988; Windfuhr, 2006, 2009).

To the best of my knowledge, there is a lack of substantive studies addressing the issues of language identification with regard to folk perspectives of what constitutes and differentiates a language in Iran (Eppler & Benedikt, 2017 is an important albeit modest exception to this gap). In other words, prior research has organized and presented linguistic data with regard to specialists' views but has not investigated folk definitions and differentiation of different types of language varieties spoken in Iran and the parameters associated with these perspectives.

In keeping with research trends in sociolinguistics, language documentation, and other language-related fields around the globe, the field of Iranian linguistics stands to benefit from language-related knowledge that proceeds from both experts and non-experts. To this end, acknowledging the crucial contribution of folk linguistic perspectives to understanding the language situation in Iran, this dissertation underlines the urgency not only to investigate linguistic perspectives, specifically those of folk, but also the linguistic and non-linguistic parameters associated with such perspectives. These parameters, which have been identified in ethnographic

and dialectological research in other parts of the world, have rarely featured in the works of Iranian and non-Iranian scholars who have examined language varieties spoken in Iran.

2.2. Conceptual Framework: Theoretical triangulation

To investigate Iranian folk perspectives about language varieties, the initial step was to determine relevant theories to establish the theoretical framework for this research project. Acknowledging that a theoretical perspective acts as a philosophical stance informing the methodology (Neuman & Robson, 2012), to determine the theoretical framework of this study, I realized that no single approach could thoroughly and systematically address the issues of language identification in Iran. This use of multiple theories in the same study known as theoretical triangulation (Denzin, 1970) helped looking at the problem at hand using multiple lenses.

This research project adopts an eclectic approach to identifying languages. I chose three theories to bear on a common problem. These theories include a folk linguistic approach, an ethnographic approach via the component of communication model and a functional approach via the language identification criteria. I use these theories as my epistemology but consider them to be more than a simple theory of knowledge. Research which analyzes folk linguistic perspectives is significant beyond addressing the existing gap in the literature regarding the issues of language identification in Iran. Such research can help scholars move towards creating more comprehensive documentation and classifications of language varieties inclusive of speech community members' perspectives. In other words, the power for making significant decisions about definition, differentiation, documentation, and organization of language varieties in language taxonomies will be shared with the members of the dominant, standard Farsi-speaking language community.

I believe that people who speak the language varieties should have the right to think about the languages the way they want, voice their thoughts with no limits imposed on them, and more importantly, be heard by scholars who document their varieties. Working with folk theory, communication model and language identification criteria, I can achieve what is of the utmost importance to me: understanding and taking seriously folk perspectives about language varieties (whether or not they have technical expertise) and acknowledging that these perspectives are relevant in preserving and maintaining the languages of my native country and the sociocultural norms associated with them.

Adopting a folk approach recommended and pursued by Hoenigswald's study (1966), to develop an understanding of and examine folk perspectives about language varieties, I began the qualitative phase of this research project with interviews. Following Kvale's (1996) typology of questions, the semi-structured interview questions were informed by Saville-Troike's (2003) communication model. Moving beyond a structural approach which identifies language varieties detached from the context of use and merely with respect to their linguistic features, this model encourages investigation of languages in association with sociocultural parameters which influence patterns of language use and linguistic perspectives.

Elaborating on the most salient social and cultural parameters, this model underlines variation in the type and number of parameters common across speech communities. The lack of a common set of parameters is based on the premise that how people identify and use language varieties varies depending on their purpose. These parameters include but are not limited to sociocultural values, speech community membership, status of language varieties within

education, social media and government, level of communication, and language use patterns based on parameters such as age, sex, social status, occupation, educational level, rural or urban residence, and belief system. Informed by these parameters, in collaboration with a panel of experts, I developed a series of questions (see Appendix A) with which to conduct the semi-structured interviews.

Before finalizing these questions, in light of Eberhard's et al. (2021) language identification criteria, I added questions related to whether measures of inherent mutual intelligibility between speakers of speech communities meaningfully influence folk linguistic perspectives. While Saville-Troike' (2003) model underlines the prominence of influential non-linguistic parameters in language identification, Eberhard et al. (2021) determine mutual intelligibility as a more fundamental indicator of whether a language variety should be identified as a distinct language or a dialect of another language. Similar to Saville-Troike's model, language identification criteria maintain that definition and differentiation of different types of language varieties are not strictly dependent on their linguistic features. Although the respective theories address the complex relationship between language varieties, not all scholars use the same set of criteria to distinguish the 'level of divergence' between a *zabān* and a *guyesh*.

Adopting an eclectic approach in the qualitative phase and based on the interview results and experts' views, I developed and validated a language identification questionnaire. In the preliminary section of this questionnaire, considering the significance of demographic information in language identification (Saville-Troike, 2003), participants were asked to provide information regarding age, gender, formal linguistic knowledge, mother tongue, city/country of birth (see

Appendix B). In the main section of the questionnaire, participants provided their linguistic perspectives on the parameters deemed influential in the theoretical framework of this research project. That is, using Likert-scale options, participants shared their perspectives about whether geography (urban vs. rural), number of speakers, status of language varieties, the extent of mutual intelligibility and shared words among language varieties, and standardized writing systems are relevant for definition and differentiation of different types of language varieties.

2.3. Methodology

To investigate folk linguistic perspectives regarding the issues of language identification in Iran, this research project made use of an exploratory sequential mixed methods research design. The ‘methodological eclecticism’ (Creswell, 2015) characteristic of mixed methods research will enable me to address these issues. In what follows, details about the participants and an overview of data collection and analysis as well as instruments are laid out. Since details regarding data collection and analysis procedure are part of the steps taken to validate the language identification questionnaire, they are presented in the respective steps.

2.3.1. Participants

The sampling process is a significant step in any research project as “it is rarely practical, efficient or ethical to study whole populations” (Marshall, 1996, p. 522). Yet, there is a high degree of confusion about conditions of sampling, since there is little agreement about best approaches to sampling and the optimum sample size in qualitative, quantitative and mixed methods studies. Choosing the right sampling approach depends on the phenomenon under study, aim of the study and research questions (Marshall, 1996). This section elaborates on the research approaches

towards selecting the participants in each phase of this exploratory mixed methods research project as well as the number of participants. Regardless of the approach to selecting the participants, the participants (see Sections 2.3.1.1. to 2.3.1.4.) were from three different universities in Tehran, Iran including Shahid Rajaei Teacher Training University (SRTTU), Iran University of Science and Technology (IUST) and Islamic Azad University Central Tehran Branch (IAUCTB). Further, some of the participants during the Exploratory Factor Analysis (EFA, a statistical approach to examine the internal reliability of a measure) and Confirmatory Factor Analysis (CFA, a similar statistical approach to EFA) were from among patients in a medical clinic in Tabriz, Iran.

While I recruited the participants for the interviews and pilot iterations, my research team assisted me with data collection during the EFA and CFA phases as well as administration of the validated questionnaire. This team included four professors in applied linguistics, six graduate students in applied linguistics, two undergraduate students in Farsi studies, an M.D., and two nurses. All of the people involved in the process of this research project had given their consent via the consent forms available prior to data collection. Further, whereas during the pilot iterations the participants received the questionnaires via email, in the EFA, CFA and the last round of data collection in Step eleven, the paper and electronic copies of the questionnaires were filled out.

2.3.1.1. Phase one: Participants in semi-structured interviews

It is common among researchers to draw a representative sample with a probability sampling technique. However, Marshall (1996) argues that in complex qualitative studies selecting a true random sample requires the researchers to know the characteristics under study of the target population. Considering the lack of folk studies in the field of Iranian linguistics, there was not

sufficient literature available which might have underlined how folk address the issues of language identification. Hence, choosing a random sample representative of Iranian folk perspectives was not practical. Further, recognizing that some participants are more likely to be knowledgeable about an unknown phenomenon under study, choosing participants at random was not necessarily more productive than selecting them by using a non-probability sampling technique (Marshall, 1996). For these reasons, I selected the participants by drawing on purposive and convenience sampling techniques from the category of non-probability sampling.

Considering the challenges at the time of conducting the qualitative phase, especially the lack of time and access to Iranian participants, I turned to convenience sampling which meant selecting the most accessible participants (Marshall, 1996). Additionally, purposeful sampling helped choose participants willing to talk for extended periods of time and able to cover a range of points of view (Rubin & Rubin, 1995). This, in turn, helped highlight similarities and differences among the participants' perspectives via access to a richer amount of data compared with probability sampling techniques which attempt to generalize (i.e., statistical inferences) from the sample being studied to the wider population of interest (Shaw, 1999).

Drawing on the conceptual framework of this study, that is, with considering a continuum of folk in particular, I decided to select participants with varying degrees of 'folkness' along this continuum. Further, although the proportion of specified characteristics of the population of interest such as their age or gender was not significant (as is the case in quota sampling), I decided to select participants with different demographic parameters as these parameters are underlined as influential in language identification. In doing so, I purposively selected (purposeful sampling)

people with no formal linguistic knowledge as well as people who had some linguistic knowledge due to taking courses in linguistics or personal interest. Lastly, these participants were from different mother tongues, age groups, level of education, and gender.

I conducted semi-structured interviews (see Appendix A) with a total number of 12 participants. As Table 3 shows, nearly 60% of the participants had taken courses in linguistics. More than half of the participants (66.7%) had BA degrees whereas 25% and 8% were MA and PhD degree holders, respectively. Data further showed that the number of participants in the age category '30-40' (33.3%) is half the number of those in the '20-30' age category (66.7%). Finally, 66.7% were female whereas 33.3% were male.

Eight participants were speakers of Farsi and four participants were speakers of Kurdish as their mother tongue. In the rest of this dissertation, I refer to these participants as Farsi and Kurdish speakers, respectively. Table 3 also shows that 50% of Farsi speakers were from Tehran and 50% were from Abadeh. The majority of these participants had taken courses in linguistics (62.5%), were BA graduates (62.5%) and in the '20-30' age category (75%). All Kurdish speakers were from Ilam. Nearly 50% of them had formal linguistic knowledge whereas 50% did not have formal linguistic knowledge. Of the Kurdish participants 75% had BA degrees. An equal number (50%) of participants in this group were in the age category '20-30' and '30-40'. Lastly, 75% of Kurdish speakers were male.

Table 3. Participants' demographic information in the qualitative phase

Mother tongue		Speech community members				Formal linguistic knowledge			Education				Age			Gender		
		Tehran	Ilam	Abadeh	Total	Yes	No	Total	BA	MA	PhD	Total	20-30	30-40	Total	Male	Female	Total
Farsi		4	0	4	8	5	3	8	5	2	1	8	6	2	8	1	7	8
		50.0%	0.0%	50.0%	100.0%	62.5%	37.5%	100.0%	62.5%	25.0%	12.5%	100.0%	75.0%	25.0%	100.0%	12.5%	87.5%	100.0%
	Kurdish	0	4	0	4	2	2	4	3	1	0	4	2	2	4	3	1	4
Total		4	4	4	12	7	5	12	8	3	1	12	8	4	12	4	8	12
		33.3%	33.3%	33.3%	100.0%	58.3%	41.7%	100.0%	66.7%	25.0%	8.3%	100.0%	66.7%	33.3%	100.0%	33.3%	66.7%	100.0%

2.3.1.2. Phase two: Participants in three pilot iterations

Following the same sampling techniques as the qualitative phase, 30, 60 and 70 participants respectively completed the questionnaire during three pilot iterations (see Steps seven and nine for more details).

Participants in the first pilot iteration

I ran the first pilot with 30 participants. As Table 4 indicates, the majority (93.3%) of these participants spoke Farsi and 6.7% spoke Kurdish as their mother tongues. In the rest of this dissertation, I refer to them as Farsi and Kurdish speakers, respectively. Nearly 57% of these participants had taken courses in linguistics while 43% of them had no formal linguistic knowledge. The majority (67%) had a BA while 20% and 13% had college and MA degrees, respectively. In terms of age, 37% were in '20-30', 33% in '30-40' and 30% in '40-50' age categories (see Table 4).

Table 4. Participants' demographic information in the first pilot iteration

Mother tongue		Farsi speech community members				Kurdish speech community members		Formal linguistic knowledge			Education			Age			Gender					
		Tehran	Kermanshah	Shusaz	Abadeh	Total	Sarandaj	Total	Yes	No	Total	College	BA	MA	Total	20-30	30-40	40-50	Total	Male	Female	Total
Farsi		13	6	5	4	28	0	0	17	11	28	6	18	4	28	10	9	9	28	11	17	28
		46.4%	21.4%	17.9%	14.3%	100.0%	0.0%	0.0%	60.7%	39.3%	100.0%	21.4%	64.3%	14.3%	100.0%	35.7%	32.1%	32.1%	100.0%	39.3%	60.7%	100.0%
	Kurdish	0	0	0	0	0	2	2	0	2	2	0	2	0	2	1	1	0	2	2	0	2
Total		13	6	5	4	28	2	2	17	13	30	6	20	4	30	11	10	9	30	13	17	30
		46.4%	21.4%	17.9%	14.3%	100.0%	100.0%	100.0%	56.7%	43.3%	100.0%	20.0%	66.7%	13.3%	100.0%	36.7%	33.3%	30.0%	100.0%	43.3%	56.7%	100.0%

Of the Farsi speakers 47% were from Tehran and the rest were from Kermanshah, Shiraz and Abadeh, respectively (see Table 4). Of these, 60.7% had taken linguistics courses. Most (64.3%) of these participants had BA, 21.4% had college and 14.3% had MA degrees. Approximately 36% were in the ‘20-30’ age category and an equal population of 32% were in the two age categories ‘30-40’ and ‘40-50’. Lastly, nearly 61% of these participants were female. Only two of the participants were Kurdish speakers from Sanandaj. They were all male BA graduates with no formal linguistic knowledge and in the two age categories ‘20-30’ and ‘30-40’.

Participants in the second pilot iteration

In the second pilot iteration, the participants spoke one of the following five language varieties as their mother tongue. These varieties included Farsi (41.7%), Turkic (20%), Kurdish (16.7%), Mazandarani (15%), and Lori (6.7%). In the rest of this dissertation, these participants are referred to as Farsi, Turkic, Kurdish, Mazandarani, and Lori speakers, respectively. Almost 85% of the participants had no formal linguistic knowledge. Nearly 55% had a BA, followed by MA/PhD (20%), high school (15%) and some (10%) were college graduates. About 52% were in the age category ‘20-30’ whereas an average of 16% were in all other age categories. Lastly, 62% of these participants were male and 38% were female (see Table 5).

Table 5. Participants’ demographic information in the second pilot iteration

		Formal linguistic knowledge			Education				Age					Gender			
		Yes	No	Total	High school	College	BA	MA/PhD	Total	Under 20	20-30	30-40	40-50	Total	Male	Female	Total
Mother tongue	Farsi	2	23	25	8	3	10	4	25	10	4	5	6	25	8	17	25
		8.0%	92.0%	100.0%	32.0%	12.0%	40.0%	16.0%	100.0%	40.0%	16.0%	20.0%	24.0%	100.0%	32.0%	68.0%	100.0%
	Kurdish	0	10	10	1	3	4	2	10	1	6	0	3	10	5	5	10
		0.0%	100.0%	100.0%	10.0%	30.0%	40.0%	20.0%	100.0%	10.0%	60.0%	0.0%	30.0%	100.0%	50.0%	50.0%	100.0%
	Turkic	4	8	12	0	0	8	4	12	0	12	0	0	12	5	7	12
		33.3%	66.7%	100.0%	0.0%	0.0%	66.7%	33.3%	100.0%	0.0%	100.0%	0.0%	0.0%	100.0%	41.7%	58.3%	100.0%
	Mazandarani	3	6	9	0	0	9	0	9	0	9	0	0	9	3	6	9
	33.3%	66.7%	100.0%	0.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	0.0%	100.0%	33.3%	66.7%	100.0%	
Lon	0	4	4	0	0	2	2	4	0	0	4	0	4	2	2	4	
	0.0%	100.0%	100.0%	0.0%	0.0%	50.0%	50.0%	100.0%	0.0%	0.0%	100.0%	0.0%	100.0%	50.0%	50.0%	100.0%	
Total	9	51	60	9	6	33	12	60	11	31	9	9	60	23	37	60	
	15.0%	85.0%	100.0%	15.0%	10.0%	55.0%	20.0%	100.0%	18.3%	51.7%	15.0%	15.0%	100.0%	38.3%	61.7%	100.0%	

Except for the two Farsi speakers who were from Tehran (see Table 79 in Appendix C), the rest of Farsi-speaking participants had no formal linguistic knowledge. Farsi speakers were mostly BA (40%) and high school (32%) graduates; 40% of these participants were under 20 years old; and 68% were female and 32% were male (see Table 5).

None of the Kurdish speakers had taken courses in linguistics. Similar to the Farsi speakers, 40% of Kurdish participants had BA degrees followed by college (30%) and MA/PhD (20%) degree holders. Of Kurdish speakers 60% were in '20-30' and 30% in '40-50' age categories. An equal number (50%) of Kurdish participants were male and female (see Table 5).

Compared with Farsi and Kurdish speakers, a lower number of Turkic and Mazandarani speakers (67%) did not have formal knowledge of linguistics; 67% of Turkic and all Mazandarani participants were BA graduates in the age category '20-30'; and 58% of Turkic and 67% of Mazandarani participants were female (see Table 5).

Lastly, none of Lori participants had taken courses in linguistics and all of them were in the age category '30-40'. Fifty percent of these participants were BA and 50% were MA/PhD graduates. Lastly, like Kurdish participants, an equal number (50%) of these participants were male and female (see Table 5).

As Table 6 shows, 56% of Farsi speakers were from Tehran (28%) and Bushehr (28%) followed by an equal population of 16% from Shiraz and Zahedan. Kurdish speakers were from Ilam (60%) and Mahabad (40%). An equal number (50%) of Turkic speakers were from Tabriz

and Zanjan. Lastly, Mazandarani speakers were from Sari (55.6%) and Amol (44.4%). Details about these speech community members are presented in Appendix C.

Table 6. Speech community members in the second pilot iteration

Farsi speech community members	Tehran	7	28.0%
	Kermanshah	3	12.0%
	Bushehr	7	28.0%
	Shiraz	4	16.0%
	Zahedan	4	16.0%
	Total	25	100.0%
Kurdish speech community members	Ilam	6	60.0%
	Mahabad	4	40.0%
	Total	10	100.0%
Turkic speech community members	Tabriz	6	50.0%
	Zanjan	6	50.0%
	Total	12	100.0%
Mazandarani speech community members	Sari	5	55.6%
	Amol	4	44.4%
	Total	9	100.0%
Lori speech community members	Khorramabad	4	100.0%
	Total	4	100.0%

Participants in the third pilot iteration

In the third pilot iteration, the majority (60%) of participants spoke Farsi as their mother tongue. This population is followed by speakers of Turkic (17%), Kurdish (14%), Gilaki (6%), and Lori (3%) language varieties. In the rest of this dissertation, these participants are referred to as Farsi, Turkic, Kurdish, Gilaki, and Lori speakers, respectively. Similar to the first pilot iteration, the majority of participants (78.6%) had no formal linguistic knowledge in this iteration; 88.6% of the participants were BA graduates, 68.6% were in the '20-30' age category and 61.4% were female (see Table 7).

Table 7. Participants' demographic information in the third pilot iteration

Mother tongue		Formal linguistic knowledge			Education				Age				Education			
		Yes	No	Total	College	BA	MA/PhD	Total	Under 20	20-30	30-40	Total	College	BA	MA/PhD	Total
Farsi	10	32	42	2	34	6	42	8	26	8	42	2	34	6	42	
	23.8%	76.2%	100.0%	4.8%	81.0%	14.3%	100.0%	19.0%	61.9%	19.0%	100.0%	4.8%	81.0%	14.3%	100.0%	
	1	9	10	0	10	0	10	2	8	0	10	0	10	0	10	
	10.0%	90.0%	100.0%	0.0%	100.0%	0.0%	100.0%	20.0%	80.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	
	3	9	12	0	12	0	12	2	10	0	12	0	12	0	12	
	25.0%	75.0%	100.0%	0.0%	100.0%	0.0%	100.0%	16.7%	83.3%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	
	0	4	4	0	4	0	4	2	2	0	4	0	4	0	4	
	0.0%	100.0%	100.0%	0.0%	100.0%	0.0%	100.0%	50.0%	50.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	
	1	1	2	0	2	0	2	0	2	0	2	0	2	0	2	
	50.0%	50.0%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	
Total	15	55	70	2	62	6	70	14	48	8	70	2	62	6	70	
	21.4%	78.6%	100.0%	2.9%	88.6%	8.6%	100.0%	20.0%	68.6%	11.4%	100.0%	2.9%	88.6%	8.6%	100.0%	

Table 7 further shows that all of Gilaki, 90% of Kurdish, nearly 75% of Farsi and Turkic, and 50% of Lori speakers had no formal linguistic knowledge. Most of the Farsi (81%) and all of Kurdish, Turkic, Gilaki, and Lori speakers had BA degrees. All of the Lori speakers, an average of 81.5% of Kurdish and Turkic, 62% of Farsi, and 50% of Gilaki speakers were in the '20-30' age category. Nearly 50% of Gilaki, 19.5% of Farsi and Kurdish, and 17% of Turkic speakers were under 20 years old. In terms of gender, all Gilaki and Lori, 67% of Turkic, 60% of Farsi, and 40% of Kurdish speakers were female.

As Table 8 shows, most Farsi speakers (71.4%) were from Tehran followed by Mashhad (12%), Yazd (7%), and an equal population of 4.8% from Semnan and Kerman. 50% of Kurdish speakers were from Ilam, 40% from Urmia and 10% from Abadan. More than half (58.3%) of Turkic speakers were from Tabriz followed by an equal population of 16.7% from Arak and Meshginshahr and 8.3% from Salmas. All Lori and Gilaki speakers were from Borujerd and Rasht, respectively. Further details regarding the speech community members in the third pilot iteration are presented in Appendix C.

Table 8. Speech community members in the third pilot iteration

Farsi speech community members	Tehran	30	71.4%
	Yazd	3	7.1%
	Mashhad	5	11.9%
	Semnan	2	4.8%
	Kerman	2	4.8%
	Total	42	100.0%
Kurdish speech community members	Abdanan	1	10.0%
	Urmia	4	40.0%
	Ilam	5	50.0%
	Total	10	100.0%
Turkic speech community members	Tabriz	7	58.3%
	Arak	2	16.7%
	Salmas	1	8.3%
	Meshginshahr	2	16.7%
	Total	12	100.0%
Lori speech community members	Borujerd	2	100.0%
	Total	2	100.0%
Gilaki speech community members	Rasht	4	100.0%
	Total	4	100.0%

2.3.1.3. Phase two: Participants in questionnaire validation process

In this phase of research, to ensure that Iranian folk with varied demographics completed the questionnaire, a team of research assistants (see Section 2.3.1.) administered it to as many speech communities as possible. With the goal of validating a contextualized questionnaire, it was filled out by 203 participants from multiple speech communities for the Exploratory Factor Analysis (EFA). For the Confirmatory Factor Analysis (CFA), 389 participants filled out the questionnaire.

Exploratory Factor Analysis (EFA) participants

The majority of EFA participants spoke Farsi (40.9%) or Turkic (34.5%) language varieties as their mother tongue. The rest were speakers of Kurdish (6.4%), Gilaki (5.4%), Mazandarani (4.4%), Lori (4.4%), or Balochi (3.9%) language varieties. In the rest of this dissertation, these participants are referred to as Farsi, Turkic, Kurdish, Gilaki, Mazandarani, Lori and Balochi speakers, respectively.

Almost 90% of these participants had no formal linguistic knowledge. A little over 50% of them had BA, 31.5% had college and 16.3% had MA/PhD degrees. The participants were dispersed across all age categories in this phase. An average of 26% were in each of the three age categories between 20 to 50 years of age, 15% were over 50 and 7% were under 20 years old. Lastly, 63% were female and 37% were male (see Table 9).

Table 9. Participants' demographic information during the Exploratory Factor Analysis (EFA)

Mother tongue		Formal linguistic knowledge			Education				Age					Gender		
		Yes	No	Total	College	BA	MA/PhD	Total	Under 20	20-30	30-40	40-50	Over 50	Total	Male	Female
Farsi	5	78	83	29	43	11	83	5	22	19	24	13	83	37	46	83
	6.0%	94.0%	100.0%	34.9%	51.8%	13.3%	100.0%	6.0%	26.5%	22.9%	28.9%	15.7%	100.0%	44.6%	55.4%	100.0%
Kurdish	3	10	13	2	8	3	13	1	3	7	2	0	13	3	10	13
	23.1%	76.9%	100.0%	15.4%	61.5%	23.1%	100.0%	7.7%	23.1%	53.8%	15.4%	0.0%	100.0%	23.1%	76.9%	100.0%
Turkic	11	59	70	24	34	12	70	2	20	17	16	15	70	27	43	70
	15.7%	84.3%	100.0%	34.3%	48.6%	17.1%	100.0%	2.9%	28.6%	24.3%	22.9%	21.4%	100.0%	38.6%	61.4%	100.0%
Mazandarani	1	8	9	2	4	3	9	2	2	3	2	0	9	3	6	9
	11.1%	88.9%	100.0%	22.2%	44.4%	33.3%	100.0%	22.2%	22.2%	33.3%	22.2%	0.0%	100.0%	33.3%	66.7%	100.0%
Balochi	0	8	8	1	6	1	8	3	2	2	1	0	8	3	5	8
	0.0%	100.0%	100.0%	12.5%	75.0%	12.5%	100.0%	37.5%	25.0%	25.0%	12.5%	0.0%	100.0%	37.5%	62.5%	100.0%
Lori	0	9	9	4	5	0	9	0	4	0	3	2	9	1	8	9
	0.0%	100.0%	100.0%	44.4%	55.6%	0.0%	100.0%	0.0%	44.4%	0.0%	33.3%	22.2%	100.0%	11.1%	88.9%	100.0%
Gilaki	1	10	11	2	6	3	11	1	5	2	2	1	11	1	10	11
	9.1%	90.9%	100.0%	18.2%	54.5%	27.3%	100.0%	9.1%	45.5%	18.2%	18.2%	9.1%	100.0%	9.1%	90.9%	100.0%
Total	21	182	203	64	106	33	203	14	58	50	50	31	203	75	128	203
	10.3%	89.7%	100.0%	31.5%	52.2%	16.3%	100.0%	6.9%	28.6%	24.6%	24.6%	15.3%	100.0%	36.9%	63.1%	100.0%

Table 9 further shows that none of Balochi and Lori, 6% of Farsi, an average of 10% of Gilaki and Mazandarani, 16% of Turkic, and 23% of Kurdish speakers had taken linguistics courses. Across all speech communities, Farsi, Turkic and Lori speakers were primarily BA and college degree holders. Kurdish, Mazandarani and Gilaki speakers mostly had BA and MA/PhD degrees. An equal number (12.5%) of Balochi speakers had college and MA/PhD degrees while the majority were BA degree holders.

An average of 26% of Farsi speakers were in the three age categories between 20 to 50; 54% of Kurdish speakers were 30 to 40 years old. Turkic speakers were dispersed across the four age categories from 20 to over 50. Mazandarani speakers were primarily (33%) between 30 to 40

years old followed by an equal population of 22.2% dispersed among all other age categories; 37.5% of Balochi speakers were under 20 and 50% were in the two age categories from 20 to 40 years of age. 44.4% of Lori speakers were in their 20s and 33.3% of them were in their 40s. Lastly, nearly 50% of Gilaki speakers were in the '20-30' age category. Regarding gender, 55.5% and 44.5% of Farsi speakers were female and male, respectively. The majority of the other speakers, however, were female (see Table 9).

Among speakers of Farsi (see Table 10), the majority were from Tehran (32.5%), followed by Isfahan (15.7%), Shiraz (14.5%), and Mashhad; 61.5% of Kurdish speakers were from Ilam. More than half (54.3%) of Turkic speakers were from Tabriz. Majority (78%) of Mazandarani and Lori speakers were from Amol and Yasuj, respectively; 62.5% of Balochi speakers were from Iranshahr and 45.5% of Gilaki speakers were from Rasht. Details regarding these speech community members are presented in Appendix C.

Confirmatory Factor Analysis (CFA) participants

Most of the participants during the CFA were speakers of Farsi (35.7%) and Turkic (30.3%), followed by speakers of Mazandarani (20%) and Kurdish (14%) language varieties. In the rest of this dissertation, these participants are referred to as Farsi, Turkic, Mazandarani, and Kurdish speakers. Similar to other phases of this research project, the majority participants (82%) had no formal linguistic knowledge. Approximately 50% had BA, 30% had college and 20% had MA/PhD degrees; 31.4% of these participants were in '20-30', an equal population of 25.4% were in '30-40' and '40-50', 11.8% were in 'over 50', and 5.9% were in 'under 20' age categories. Nearly 61% were female and 39% were male (see Table 11).

Table 10. Speech community members during the Exploratory Factor Analysis (EFA)

Farsi speech community members	Tehran	27	32.5%
	Shiraz	12	14.5%
	Isfahan	13	15.7%
	Mashhad	10	12.0%
	Kermanshah	8	9.6%
	Kerman	5	6.0%
	Bushehr	4	4.8%
	Yazd	2	2.4%
Kurdish speech community members	Sanandaj	5	38.5%
	Ilan	8	61.5%
	Total	13	100.0%
Turkic speech community members	Tabriz	38	54.3%
	Ardabil	10	14.3%
	Urmia	12	17.1%
	Tarom	10	14.3%
Mazandarani speech community members	Amol	7	77.8%
	Babol	2	22.2%
	Total	9	100.0%
Balochi speech community members	Iranshahr	5	62.5%
	Chabahar	3	37.5%
	Total	8	100.0%
Lori speech community members	Yasuj	7	77.8%
	Khorramabad	2	22.2%
	Total	9	100.0%
Gilaki speech community members	Rasht	5	45.5%
	Lahijan	4	36.4%
	Fuman	2	18.2%
	Total	11	100.0%

Table 11. Participants' demographic information during Confirmatory Factor Analysis (CFA)

		Formal linguistic knowledge			Education				Age					Gender			
		Yes	No	Total	College	BA	MA/PhD	Total	Under 20	20-30	30-40	40-50	Over 50	Total	Male	Female	Total
Mother tongue	Farsi	17	122	139	47	70	22	139	7	45	27	41	19	139	65	74	139
		12.2%	87.8%	100.0%	33.8%	50.4%	15.8%	100.0%	5.0%	32.4%	19.4%	29.5%	13.7%	100.0%	46.8%	53.2%	100.0%
	Kurdish	14	40	54	19	25	10	54	3	17	17	13	4	54	20	34	54
		25.9%	74.1%	100.0%	35.2%	46.3%	18.5%	100.0%	5.6%	31.5%	31.5%	24.1%	7.4%	100.0%	37.0%	63.0%	100.0%
	Turkic	30	88	118	34	57	27	118	4	35	32	31	16	118	48	70	118
		25.4%	74.6%	100.0%	28.8%	48.3%	22.9%	100.0%	3.4%	29.7%	27.1%	26.3%	13.6%	100.0%	40.7%	59.3%	100.0%
Mazandarani	9	69	78	20	39	19	78	9	25	23	14	7	78	20	58	78	
		11.5%	88.5%	100.0%	25.6%	50.0%	24.4%	100.0%	11.5%	32.1%	29.5%	17.9%	9.0%	100.0%	25.6%	74.4%	100.0%
	Total	70	319	389	120	191	78	389	23	122	99	99	46	389	153	236	389
	18.0%	82.0%	100.0%	30.8%	49.1%	20.1%	100.0%	5.9%	31.4%	25.4%	25.4%	11.8%	100.0%	39.3%	60.7%	100.0%	

Furthermore, Table 11 shows that an average of 12% of Farsi and Mazandarani as well as an average of 25.5% of Kurdish and Turkic speakers had no formal knowledge of linguistics. An average of 49% of Farsi, Kurdish, Turkic, and Mazandarani speakers had BA degrees. The most populated age category across all groups of speakers was '20-30'. Regarding this, 32.5% of Farsi speakers, 31.5% of Kurdish, 29.7% of Turkic, and 32.1% of Mazandarani speakers were in the

respective age category. In terms of gender, among all groups, Mazandarani speakers had the highest population (74.5%) of female speakers followed by Kurdish (63%), Turkic (59.3%), and Farsi (53.2%) speakers.

As Table 12 shows, 45.3% of the population of Farsi speakers were from Tehran. 39% and 24% of Kurdish speakers were from Mahabad and Sanandaj, respectively. The majority of Turkic speakers (68.7%) were from Tabriz. Most of Mazandarani speakers were from Behshahr (24.4%), Sari (19.2%), Nowshahr (17.9%), and Babol (14.1%). Further details about these speech community members are provided in Appendix C.

Table 12. Speech community members during Confirmatory Factor Analysis (CFA)

Farsi speech community members	Tehran	63	45.3%
	Karaj	10	7.2%
	Qom	7	5.0%
	Ahvaz	5	3.6%
	Bushehr	10	7.2%
	Abadeh	5	3.6%
	Shiraz	7	5.0%
	Isfahan	11	7.9%
	Zahedan	2	1.4%
	Borazjan	2	1.4%
	Kerman	3	2.2%
	Kemanshah	5	3.6%
	Semnan	5	3.6%
	Yazd	4	2.9%
Total	139	100.0%	
Kurdish speech community members	Ilam	9	16.7%
	Sanandaj	13	24.1%
	Baneh	3	5.6%
	Paveh	4	7.4%
	Mahabad	21	38.9%
	Abdanan	4	7.4%
Total	54	100.0%	
Turkic speech community members	Tabriz	80	67.8%
	Urmia	15	12.7%
	Ardabil	10	8.5%
	Zanjan	13	11.0%
	Total	118	100.0%
Mazandarani speech community members	Nowshahr	14	17.9%
	Amol	10	12.8%
	Babol	11	14.1%
	Sari	15	19.2%
	Behshahr	19	24.4%
	Babolsar	9	11.5%
	Total	78	100.0%

2.3.1.4. Phase three: Participants in administration of the questionnaire

In this phase, considering the challenges of a global pandemic, distribution of the questionnaire was limited to the participants accessible for the research. Thus, in phase three, 192 participants completed the language identification questionnaire. As summarized in Table 13, the majority of participants were speakers of Turkic (37.5%) and Kurdish (36.5%), followed by speakers of Farsi (19.8%) and Mazandarani (6.3%) language varieties. In the rest of this dissertation, these participants are referred to as Turkic, Kurdish, Farsi, and Mazandarani speakers, respectively.

Table 13. Participants' demographic information in questionnaire administration

Mother tongue		Formal linguistic knowledge			Education				Age					Gender			
		Yes	No	Total	College	BA	MA/PhD	Total	Under 20	20-30	30-40	40-50	Over 50	Total	Male	Female	Total
Farsi		9	29	38	11	15	12	38	3	16	10	7	2	38	18	20	38
		23.7%	76.3%	100.0%	28.9%	39.5%	31.6%	100.0%	7.9%	42.1%	26.3%	18.4%	5.3%	100.0%	47.4%	52.6%	100.0%
Kurdish		13	57	70	35	17	18	70	4	18	34	14	0	70	24	46	70
		18.6%	81.4%	100.0%	50.0%	24.3%	25.7%	100.0%	5.7%	25.7%	48.6%	20.0%	0.0%	100.0%	34.3%	65.7%	100.0%
Turkic		16	56	72	20	32	20	72	15	31	15	8	3	72	26	46	72
		22.2%	77.8%	100.0%	27.8%	44.4%	27.8%	100.0%	20.8%	43.1%	20.8%	11.1%	4.2%	100.0%	36.1%	63.9%	100.0%
Mazandarani		0	12	12	0	11	1	12	4	5	2	0	1	12	1	11	12
		0.0%	100.0%	100.0%	0.0%	91.7%	8.3%	100.0%	33.3%	41.7%	16.7%	0.0%	8.3%	100.0%	8.3%	91.7%	100.0%
Total		38	154	192	66	75	51	192	26	70	61	29	6	192	69	123	192
		19.8%	80.2%	100.0%	34.4%	39.1%	26.6%	100.0%	13.5%	36.5%	31.8%	15.1%	3.1%	100.0%	35.9%	64.1%	100.0%

From among the 192 participants, 80.2% had not taken courses in linguistics. In terms of educational background, 34.4% had a college, 39.1% had a BA and 26.6% of the participants had a graduate degree. In terms of age, 36.5% and 32% of the participants were in the '20-30' and '30-40' age categories, respectively; smaller proportions of 13.5% were under 20 and 15% were over 50 years old. Lastly, 64% of the participants were female while 36% were male (see Table 13).

Table 13 further shows that Mazandarani speakers were the only group none of whom had formal education in linguistics. Farsi (39.5%), Turkic (44.4%) and Mazandarani (91.7%) speakers were primarily BA degree holders whereas 50% of Kurdish had college degrees. A little over 40%

of Farsi, Turkic and Mazandarani speakers were in the age category '20-30' while nearly 49% of Kurdish speakers were in their 30s. In all groups, the number of female speakers was more than 50% of the sample population. Farsi speakers had the smallest female population (52.6%) whereas Mazandarani speakers had the largest one (91.7%).

Table 14 shows that 26.4% of Farsi speakers were from Tehran followed by Mashhad, Kermanshah and Bushehr. Among Kurdish speakers, 24.3% were from Urmia, an equal population of 21.4% were from Ilam and Sanandaj, and 18.6% were from Kermanshah. More than 50% of Turkic speakers were from Tabriz. Nearly 33.3% of Mazandarani speakers were from Amol and 50% were from Sari and Babol. Further details about these speech community members are presented in Appendix C.

Table 14. Speech community members in the administration of language identification questionnaire

Farsi speech community members	Tehran	10	26.3%
	Mashhad	4	10.5%
	Isfahan	3	7.9%
	Shiraz	3	7.9%
	Kermanshah	4	10.5%
	Ghom	2	5.3%
	Ahvaz	3	7.9%
	Abadeh	2	5.3%
	Bushehr	4	10.5%
	Kerman	3	7.9%
	Total	38	100.0%
Kurdish speech community members	Ilam	15	21.4%
	Kermanshah	13	18.6%
	Sanandaj	15	21.4%
	Mahabad	6	8.6%
	Urmia	17	24.3%
	Sardasht	4	5.7%
Total	70	100.0%	
Turkic speech community members	Tabriz	37	51.4%
	Urmia	11	15.3%
	Zanjan	9	12.5%
	Ghazvin	7	9.7%
	Ardabil	8	11.1%
Total	72	100.0%	
Mazandarani speech community members	Amol	4	33.3%
	Sari	3	25.0%
	Babol	3	25.0%
	Tonekabon	2	16.7%
Total	12	100.0%	

2.3.2. Overview of data collection and analysis

This research project started with a qualitative phase during which I conducted semi-structured interviews. The purpose of this phase was to develop tentative components of the language identification questionnaire with respect to the main themes raised during semi-structured interviews (see Section 3.2.4.). I structured the interviews according to Kvale's (1996) typology of question types (see Section 3.2.2.). That is, I adopted ten question types, namely: introductory, follow-up, probing, specifying, direct, indirect, structuring, interpreting, and

throwaway questions. I practiced periods of silence in between questions, as the last part of this typology recommended by Kvale.

Following the qualitative data analysis and having reviewed the related literature, I developed a questionnaire (see Appendix B). To ensure that the generated questionnaire is both reliable and valid, I further piloted it in three iterations (see Section 4.5. for pilot one and Section 4.7. for pilots two and three) and sought the expertise of an expert panel of linguists and sociolinguists (see Section 3.1.). In light of the collected pilot data, the expert panel helped modify this questionnaire in terms of its content, language and structure (see Section 4.4.).

As the last step, having run Exploratory Factor Analysis (EFA), I validated this questionnaire via Confirmatory Factor Analysis (CFA) and tested its fitness via the collected data (see Section 4.8.). I administered this validated questionnaire among Iranian folk to investigate their perspectives on language identification in Iran (see Section 5.1.). Details regarding data collection and analysis procedure in each of these phases are presented in relevant sections.

For clarity purposes, the coding procedure of the questionnaire items is provided in this section here. Likert-scale responses included ‘don’t know’=1, ‘disagree’=2, ‘somewhat disagree’=3, ‘somewhat agree’=4, and ‘agree’=5. For sake of data analysis, Likert-scale responses ‘agree’ and ‘partly agree’ as well as ‘disagree’ and ‘partly disagree’ were combined. At this point of the research, interpretation of ‘partly agree’ and ‘partly disagree’ as separate categories was not feasible because the reasons behind people’s responses with regard to these Likert-scale categories were not clear.

Note that all of the Likert-scale tables are based on the five Likert-scale options. Although I believed that the participants must be given uncertainty of choice via ‘don’t know’, due to lack of data as to why this option was selected from among the available Likert-scale options, I only commented on participants’ extent of agreement/disagreement in every component of the questionnaire. Lastly, considering the range of perspectives about the use of linguistic terms, a glossary provides a list of Farsi linguistic terms and their English equivalents used in this study (see p. XXIII).

3. Hybrid thesis: Part two (Phase one)

In studies with no existing measures to assess phenomena under study, the construction of measuring devices is the most important segment (Shultz et al., 2014). These devices need to be accurate measurements of constructs which are representations of behaviours that are unobservable and difficult to measure (Neuman & Robson, 2012). On the significance of sound measurement tools, Schoenfeldt (1984) states that “many well-conceived research studies have never seen the light of day because of flawed measures” (p. 78). Hence, the use of a well-established framework which guides researchers through multiple stages of scale development prevents the creation of faulty tools (Hinkin, 1998). Part two of this dissertation describes and discusses the development of a measurement tool in three phases comprising eleven steps. The scale development process was informed by Hinkin (1998), Neuman and Robson (2012) and Shultz et al. (2014).

As mentioned earlier in this dissertation, to address the two research questions of this project, phase one was a qualitative exploration of folk linguistic perspectives and the associated parameters. I took two steps to complete this phase. In **Step one**, I convened a panel of linguists and sociolinguists with relevant training and experience, partly familiar with the topic under study (i.e., language identification) and/or development of measures for use in questionnaires. This panel looked over the development of this questionnaire in all three phases and eleven steps of this research project including item generation, development of the interviews as well as questionnaire demographic questions, internal consistency, and face and content validity. Details regarding this panel’s discussions and their extent of involvement are included in the respective steps. In **Step two**, to determine the construct to be assessed with the questionnaire, I consulted the expert panel,

conducted a literature review, and held semi-structured interviews. The semi-structured interviews underlined language identification to be a multidimensional construct with multiple related components. Results obtained from this phase were presented in 2017 at the Third Kurdish Studies Conference at the University of Exeter, UK.

3.1. Phase one- Step one: Ethics and assembly of an expert panel

Initiating this research project, I understood that developing a quantitative measurement is a challenging process which requires persistence, integrity and interaction with scholars to combine outlooks and produce knowledge (Neuman & Robson, 2012). As such, to address language identification issues in a systematic way and to select the appropriate techniques to address them, I completed the ethics process through the Carleton University Research Ethics Board-A (CUREB-A) at Carleton University and convened an expert panel to help develop this questionnaire. Completing the ethics process clarified details with regard to recruitment, anonymity and integrity of participants, conducting interviews and thus completing the qualitative phase, formats (online and paper) of the questionnaire and alternatives of distribution, and written and oral consents.

With the possibility of various approaches taken by speech communities in identifying language varieties, a panel of seven linguists/sociolinguists (including the three supervisory committee members) was convened to closely evaluate the generation and accumulation of items in this measurement model. Among the panel members were three professors working on Iranian linguistics. These scholars speak Farsi and have researched specific Iranian (Iranic) language varieties inside and outside of Iran. Among these language varieties are Kurdish, Lori, Kumzari, and Gulf dialects of Farsi, to name a few. Among their numerous publications, they have addressed linguistic and sociolinguistic status of Iranian (Iranic) language varieties via structural, functional and social approaches.

In addition to the respective experts, this study benefited from the contributions of two professors with expertise in cross-cultural communication and two professors with expertise in qualitative and quantitative approaches to social research. Their feedback was specifically helpful in generating questions for the semi-structured interviews (see Appendix A), generating components and items, establishing reliability, and face and content validity of this measurement tool.

3.2. Phase one- Step two: Qualitative exploration of folk linguistic perspectives

To address issues of language identification, it would be ideal to benefit from multiple interviews and ethnographic observations as valuable tools and the central methods of collecting qualitative data. However, these are not the most efficient tools due to limitations such as time constraint and financial issues. Such limitations inherent in these methodological tools make the use of validated quantitative questionnaires with a large population sample a viable alternative.

Although the available literature (e.g., see Eberhard et al, 2021; Hymes, 1974; Preston, 1989, Saville-Troike, 2003) assisted in developing components and items, folk perspectives about language varieties have been more fully explored in North American and European countries. Therefore, it was challenging to identify and generate items which represented folk perspectives about language varieties spoken in Iran. In generating items and to establish a questionnaire with applicability in the context of Iran, I adopted an ‘inductive approach’ (Hunt, 1991) to qualitative data collection using Kvale’s (1996) stages of interview investigation which helped create the preliminary items of the questionnaire.

Briefly speaking, in an inductive approach also known as classification from above, a researcher develops a measurement tool inductively by asking a sample population to describe their feelings or provide their opinions about a subject matter. This approach allows researchers to condense extensive raw data and establish a link between the research objectives and significant themes generated based on the raw data (Thomas, 2006). Although conducting interviews was not the primary method of data collection in this research project, I employed this method to gather preliminary data about folk perspectives with respect to issues of language identification. Having conducted the semi-structured interviews and transcribed and studied the data repeatedly, I adopted an eclectic approach to coding. Coded data were categorized and later conceptualized into broad themes. This process led to twelve main themes. Through the development of a summary of themes, I established an understanding of parameters significant to people in language identification. After that, items were derived from the themes and with respect to sample quotes.

Following Kvale (1996, p. 88), the interview investigation was outlined in multiple stages: thematizing, the interview itself, transcribing, analyzing, verification and reporting (see Table 15).

Table 15. Stages of an interview investigation adapted from Kvale (1996)

Interview stage	Definition
Thematizing	Formulate the purpose of the investigation and describe the concept of the topic to be investigated before the interviews start.
Designing	Plan the design of the study, taking into consideration all seven stages, before the interview starts.
Interviewing	Conduct the interviews based on an interview guide and with a reflective approach to the knowledge sought.
Transcribing	Prepare the interview material for analysis, which commonly includes a transcription from oral speech to written text.
Analyzing	Decide, based on the purpose and topic of the investigation, and on the nature of the interview material, which methods of analysis are appropriate.

Verifying	Ascertain the generalizability, reliability, and validity of the interview findings. Reliability refers to how consistent the results are, and validity means whether an interview study investigates what is intended to be investigated.
Reporting	Communicate the findings of the study and the methods applied in a form that lives up to scientific criteria, takes the ethical aspects of the investigation into consideration, and that results in a readable product.

3.2.1. Thematizing

Among my goals was to keep the spontaneity of the interview structure in search of more spontaneous answers. That said, all the interviewees had received written consent forms in which the structure, purpose and time framework of the interview as well as basic knowledge of the subject matter in this research project as a whole were explained. This gave me an initial opportunity not only to elaborate on my goals in simple language, but also remind them that these interviews were conversations during which I was interested in hearing their perspectives about the issues of language identification in Iran. Before the beginning of every interview session, the interviewee was invited to ask questions and/or share concerns. Once again, I explained the purpose of the interview.

3.2.2. Designing and interviewing

In structuring and designing the interview, I adopted Kvale's (1996, pp. 133-135) question typology. That is, every interview included ten question types (see Table 16).

Table 16. Types of interview questions with examples

Type of question	Definition	Example
Introductory	To kick start the conversation and move to the main interview	“Can you tell me about [...]?”
Follow-up	To direct questioning to what has just been said	Nodding, “mm”, Repeating significant words
Probing	To draw out more complete narratives	“Could you say something more about that?” “Can you give a more detailed description of what happened?” “Do you have further examples of this?”
Specifying	To develop more precise descriptions from general statements	“What did you think then?” “What did you actually do when you felt a mounting anxiety?”
Direct	To elicit direct responses	“Have you ever received money for good grades?” “When you mention competition, do you then think of a sportsmanlike or a destructive competition?”
Indirect	To pose projective questions	“How do you believe other pupils regard the competition of grades?”
Structuring	To refer to the use of key questions to finish off one part of the interview and open up another, or to indicate when a theme is exhausted by breaking off long irrelevant answers	“I would now like to introduce another topic [. . .]”
Silence	To allow pauses, so that the interviewees have ample time to associate and reflect, and break the silence themselves with significant information	
Interpreting	Similar to some forms of probing questions, to rephrase an interviewee’s answer to clarify and interpret rather than to explore new information	“You then mean that [. . .]?” “Is it correct that you feel that [. . .]?” “Does the expression [. . .] cover what you have just expressed?”
Throwaway	To serve a variety of purposes, i.e. to relax the subject when sensitive areas have been breached.	“Oh, I forgot to ask you [. . .]”

Following Kvale’s guide, with a focus on the conceptual framework of this research project and to answer the research questions, I incorporated a series of questions (see Table 17 for example questions) which helped direct the conversation towards the issues of language identification and seek participants’ linguistic perspectives in this regard.

Table 17. Example question type following Kvale’s (1996) typology

Type of question	Example question
Introducing	Can you tell me a bit about yourself? Where were you born and how old are you?
Follow-up	Is that your first language? Do you know any other language varieties?
Probing	Can you speak all these language varieties? How about reading and writing? Can you read and write in them?
Specifying	Which language variety do you use for your personal matters such as praying, self-talk or thinking out loud? What is the language variety that your parents chose to communicate with you?
Direct	Was it important to you/your parents that you learn Farsi too? [in case of a first language other than Farsi] Have you ever had teachers at school who did not speak the language variety that you spoke at home?
Indirect	Do you think people switch between language varieties in different occasions and under different circumstances? Do you think the speakers of two separate languages would be able to understand each other?
Structuring	Now I am going to move to another category of questions which investigates your opinion about what language varieties are. What kind of language variety would you say Farsi or Arabic is? Do you think Isfahani or Shirazi is a ... too?
Silence	
Interpreting	So ... do you mean that Farsi/Arabic are different from/similar to Shirazi or Isfahani?
Throwaway	I wanted to ask you about your feelings about the varieties that you speak. Do you have a sense of belonging to any of them?

Following Kvale’s (1996) suggestion, I used the introductory questions to establish rapport with the interviewees. Although these opening questions were not necessarily related to the subject matter at hand, they helped provide a safe context where the interviewees felt comfortable to share their ideas. Among the techniques which I practiced were clarifications with respect to the categories of the questions asked. That is, before entering every category type (e.g., demographic information or similarities and differences between language varieties) I introduced the category.

To ensure that I captured their narratives fully, I kept the flow of conversations going by asking multiple question types specifically follow-up, probing and specifying questions. The purpose was to verify my interpretations of the interviewees' responses regarding the underlying topic in the course of interviews. As Kvale (1996, p. 189) asserts, "the interviewer condenses and interprets the meaning of what the interviewee describes and 'sends' the meaning back, ideally until there is only one possible interpretation left or the multiple understandings of a theme by the subject are known".

I further tried to keep the questions brief and simple by providing examples. Depending on their demographics, some interviewees requested more clarifications and examples compared with others. This, in turn, caused some of the interviews to be longer than others. It was important that the interviewees did not feel rushed or lose confidence due to lack of knowledge in specific subject matters. Each of the interviews lasted between 60 and 90 minutes. Depending on their demographics including number of language varieties one spoke, places of residence and formal linguistic knowledge, interviewees did not answer all the questions included.

3.2.3. Transcribing and analyzing: Eclectic approach to coding

Having conducted the interviews, I listened to each of them twice to concentrate not only on the interviewees' responses, but also listen to their pauses, tone and intonation as they spoke. Then, I transcribed each interview by writing down every detail said by the interviewees and further added the notes which I had taken prior, during and after each interview session. In the data analysis, I started the coding process merely based on the collected data. I further used in-vivo coding where I used the terminology and language of the interviewees to derive codes from the

data itself. To provide an in-depth comprehension about participants' perspectives regarding language varieties, I employed values coding (Saldaña, 2013) which entails the subjective-value perspectives. For example, some of the interviewees believed that their religion and language variety are tied together. Moreover, structural coding helped with categorizing and describing the meaning of data parts which could best explain sociocultural parameters (age, gender, role-relationships, occupation, etc.) associated with folk perspectives to further examine participants' approach to utilizing linguistic labels such as *zabān*, *guyesh* and *lahjeh*. Therefore, considering the research questions, I followed an eclectic approach to coding data (Saldaña, 2013).

As a result of categorizing all the codes (not all speech communities resulted in equal number of codes) gained from each speech community, patterns were generated based on the relationship between codes and code frequencies. Categories emerged with regard to research questions (second cycle coding, Saldaña, 2013). It is noteworthy that the number and order of these categories were not similar across the speech communities as the structure of interviews allowed for omission and/or rearranging some topics depending on the individuals.

After generating the categories, the coding process continued with re-examining the transcriptions, codes and categories as defined by parameters established in the conceptual framework of this research. In this process, I observed that some of the categories could be explained using theories in the conceptual framework. The following presents the themes with regard to emergent categories and research questions. The themes generated from the speech communities helped develop the components of the questionnaire (see Table 18).

Table 18. Extracted themes and sample quotes via an eclectic approach to coding

Theme	Sample quote
Geography	<ul style="list-style-type: none"> • Whether a language variety is spoken in a small town or a big city cannot determine what it is. • A language variety spoken in villages is a <i>guyesh</i>. • A language variety spoken in cities is a <i>zabān</i>.
Number of speakers	<ul style="list-style-type: none"> • It is important to know the number of people who speak a language variety. • A language variety spoken by millions of people must be a <i>zabān</i>. • Whether a few millions or thousands speak a language variety cannot be important enough to decide the type of this language variety.
Status of language varieties	<ul style="list-style-type: none"> • <i>Zabān</i> is a variety recognized by the government. • <i>Zabān</i> is a variety which is used in the education system. • <i>Guyesh</i> is a language variety with no official status. • <i>Lahjeh</i> is a language variety with no official status. • My language variety is only occasionally used in the education system and government sectors. • The official <i>zabān</i> of the country [i.e., Farsi] is more acceptable and useful.
Mutual intelligibility	<ul style="list-style-type: none"> • Understanding is a general term which makes application of a firm borderline challenging. • Understanding is a general term which makes precise definition of language varieties challenging. • Mutual intelligibility among speakers can be low, moderate or high. • It is challenging to define a language variety when the extent of mutual intelligibility is neither low nor high. • Lower extents of mutual intelligibility between speakers means two separate <i>zabān</i>. • Lower extents of mutual intelligibility between speakers could mean two <i>guyesh</i> of the same <i>zabān</i>. • Many parameters decide whether a variety is <i>zabān</i> or <i>guyesh</i>. • Religious affiliations and sociocultural regulations decide what a variety is. • There is no difference between <i>guyesh</i> and <i>lahjeh</i>. • Two different <i>guyesh</i> have different grammar and words. • Two different <i>lahjeh</i> have different ways of pronouncing the same words. • Understanding <i>lahjeh</i> is not difficult. • Understanding <i>guyesh</i> is sometimes difficult. • Two <i>zabān</i> are not mutually intelligible. • Understanding <i>guyesh</i> depends on the extent to which I am familiar with its words. • <i>Zabān</i>, <i>guyesh</i> and <i>lahjeh</i> are part of the communities where they are used.
Writing system and extent of shared words	<ul style="list-style-type: none"> • Two <i>zabān</i> do not share many of their words. • Two <i>guyesh</i> share more of their words compared with two <i>zabān</i>. • Two <i>lahjeh</i> share more of their words compared with two <i>zabān</i>. • Two <i>lahjeh</i> which belong to the same <i>zabān</i> share almost all their words. • Every <i>zabān</i> must have a writing system. • A <i>guyesh</i> which is only mutually intelligible to lower extents has a writing

	<p>system.</p> <ul style="list-style-type: none"> • A <i>guyesh</i> which is moderately intelligible does not have a writing system.
Exposure to language varieties	<ul style="list-style-type: none"> • Exposure to other language varieties leads to familiarity with their vocabulary and grammar. • Exposure to other language varieties reinforces higher extents of mutual intelligibility.
Setting and purpose of communication	<ul style="list-style-type: none"> • Defining a variety depends on the context in which it is used. • Defining a variety depends on the purpose for which it is used. • A language variety which is used in the government is a <i>zabān</i>. • A language variety which is used for political affairs is a <i>zabān</i>. • A language variety which is used in the government and/or for political affairs is a <i>zabān</i> no matter the extent of mutual intelligibility. • A language variety which is used in the government and/or for political affairs is a <i>zabān</i> no matter the extent of shared words and/or grammar.
Demographic information	<ul style="list-style-type: none"> • Young well-educated people who know multiple language varieties can differentiate <i>zabān</i> and <i>guyesh</i>. • Illiterate people who have been exposed to their mother tongue only and lived in one city their entire life cannot define types of language varieties.
Relationships	<ul style="list-style-type: none"> • I use the official <i>zabān</i> [i.e., Farsi] in the education system and with the teachers. • I use my own language variety with my friends and classmates. • I must use the official <i>zabān</i> [i.e., Farsi] with my manager and colleagues at work. • I sometimes use my own variety with my colleagues.
Diglossia and bilingualism	<ul style="list-style-type: none"> • Switching between language varieties is for the purpose of communication outside household. • My own language variety is the dominant language variety. • I pray and express emotions in my own variety. • External parameters (e.g., migration) do not affect the vitality of my mother tongue.
Language and identity	<ul style="list-style-type: none"> • My language variety is part of who I am. • My language variety is part of my community. • Language skills such as reading and writing a variety has little to do with identifying with it. • Exposure to a second and third language variety can develop a sense of belonging to more than one language variety.
Religious affiliation	<ul style="list-style-type: none"> • My religion and my variety are tied together. • We use our language variety to maintain the unity of our religious community. • We use our language variety to separate ourselves from others [i.e., people with different religious affiliations].

3.2.4. Themes generated based on semi-structured interviews

Generated themes (see Table 18) brought to light multiple parameters which, from the interviewees' perspectives, influenced definition, differentiation and use of different types of language varieties in different situations. These qualitative results echoed the significance of the conceptual framework theories in addressing the issues of language identification.

3.2.4.1. Geography (urban vs. rural)

Nearly half of all three speech communities underlined the relevance of geography in language identification. These participants mentioned that “language varieties such as Farsi are spoken in big cities, including the capital”. The rest, however, asserted that “whether a language variety is spoken in a small town or a big city cannot determine what it is”. This could be partly due to social and linguistic heterogeneity in urban areas (Trudgill, 1974). Wolfram and Schilling (2016) further disapprove of incorrect assumptions which define dialects as the socially disfavored rural varieties spoken by socially subordinate groups of speakers.

Participants who spoke Tehrani Farsi, a colloquial norm spoken in Tehran, added that “for years, people from small towns and villages have migrated and settled down in big cities such as Tehran”. Thus, “it cannot be assumed that a language variety spoken in a city is a *zabān*”. They further explained that “if you walk down Valiasr or Enghelāb street [two of the busiest streets in Tehran], you can hear different *zabān* and *guyesh*”. These perspectives could indicate that some participants did not consider urban geography as areas where people speak only a specific language variety. In addition, these perspectives also point to cities as major language contact situations where people bring the culture and language of their origin (see Preston, 2016; Wolfram &

Schilling, 2016). However, these qualitative findings are not in line with some of the folk studies of rural and urban areas (e.g., see Cramer, 2016b; Evans, 2016) which show that, with respect to folk perspectives, the urban vs. rural dichotomy is a strong divide regardless of the linguistic variation in these areas.

3.2.4.2. Number of speakers

Nearly two-thirds of the interviewees believed that the number of people who speak a language variety can determine whether the said variety is *zabān*, *guyesh* or *lahjeh*. According to these participants, “a language variety spoken by millions of people must be a *zabān*”. Other interviewees asserted that “whether a few millions or thousands speak a language variety cannot be important enough to decide the type of this language variety”.

3.2.4.3. Status of language varieties

Folk linguistic studies (e.g., see Niedzielski & Preston, 2000; Preston, 1989) assert that people generally tend to evaluate language varieties in terms of prestige, suitability and use in various sectors of a society. By way of example, the participants specified Tehrani Farsi as “a prestigious and socially accepted variety across many speech communities in Iran”. Except for Kurdish participants, the rest explained that “*zabān* is a variety recognized by the government and used in the education system”. Some of the Kurdish and Abade’i participants believed that their variety was “only occasionally used” in education or government sectors.

Participants identified *guyesh* and *lahjeh* as varieties with no official status. While most of the participants differentiated the respective types of varieties in terms of their linguistic features including “grammar, vocabulary and pronunciation”, they recognized both as varieties “with no

official status in the country”. In the same way, scholars (e.g., see Dabir-Moghaddam, 2008; Modarresi, 2014) advocate this perspective and confirm that these varieties may not have the sociopolitical and cultural significance of Farsi in Iran, are spoken in a limited geographical territory, and have limited use in education and social media.

3.2.4.4. Mutual intelligibility

The theme ‘mutual intelligibility’ was referred to by all speech communities as a prominent factor in definition and differentiation of types of language varieties. The participants explained that “the type of a language variety depends on how much of it I understand”. Similarly, Eberhard et al. (2021) confirms that people’s approaches to defining language varieties are reliant on the degree to which a variety is understandable based on personal knowledge of one’s own variety.

The participants further identified extents of mutual intelligibility among speakers of language varieties as “low, moderate or partial and high”. They continued to say that “understanding is such a general term which makes application of a firm borderline among and precise definition of language varieties challenging goals to achieve”. As a result, it was not clear as to what extent and in what ways individuals understood language varieties.

An interesting point raised by a few of Tehrani and Abade’i participants was about a linguistic label between language varieties. They mentioned that “not all varieties are clearly a *zabān* or a *guyesh*”. Sometimes, a language variety “is neither a *zabān* nor a *guyesh*, but something in between”. These participants explained that “some varieties have some features of a *zabān* and some features of a *guyesh*”. As such, “an additional term to refer to these varieties is a good idea”.

Although mutual intelligibility was among the first parameters underlined by participants in whether two varieties should be separate *zabān* or *guyesh* of the same *zabān*, as Modarresi (2014) points out, it is only practical in circumstances where mutual intelligibility between language varieties is high or low. One such example is the Abade’i language variety. When asked to express their views about this language variety, while half of both Kurdish and Tehrani participants did not know of this variety, the other half agreed that “it is a *guyesh* of Farsi” due to high extent of understanding based on knowledge of one’s own language variety. Participants confirmed that defining and differentiating types of language varieties “is challenging and even impractical when the extent of mutual intelligibility is neither low nor high”. This lines up with what scholars (Nettle, 1999; Saville-Troike, 2003) have stated about the inadequacy of mutual intelligibility in identifying language varieties. That is, mutually understanding language varieties is insufficient proof to determine if language varieties belong to one or multiple languages. Regarding this, the majority of participants argued that “Turkic, Gilaki, and Balochi varieties should be considered separate *zabān*” because they could not understand these varieties “at all or to very limited extents”. However, the respective varieties “are considered different *guyesh* which belong to Farsi”.

For the participants of this study, a lower extent of mutual intelligibility between speakers of two language varieties was not always an indication of two separate *zabān*. For example, a few of Tehrani and Abade’i participants mentioned that in some cases “it could mean two *guyesh* of the same *zabān*”. Under such circumstances, other parameters determine whether a language variety is a *guyesh* or a *zabān*. As scholars have pointed out, identification of language varieties is

always a combination of linguistic and non-linguistic parameters including social, cultural and individual (e.g., see Eberhard et al., 2021; Saviile-Troike, 2003). Regarding this, Kurdish and Abade'i participants believed that in identifying language varieties "religious affiliations and sociocultural regulations of speech communities" are among influential parameters which can determine whether two language varieties should be considered two separate *zabān*. Considering these non-linguistic affiliations, the participants stated that "establishing a borderline among language varieties and applying linguistic labels such as *zabān*, *guyesh* and *lahjeh* are challenging".

Most of the participants elaborated on differences between types of language varieties despite challenges in language identification and recurrent dismissal of people's linguistic views as naïve misunderstandings of language (e.g., see Niedzielski & Preston, 2000; Paveau, 2011; Preston, 1989; Wilton & Stegu, 2011). Three of the participants maintained that "there is no difference between *guyesh* and *lahjeh*" or mentioned that they were not "aware of any differences between the two". The majority of participants, however, defined two *guyesh* as "varieties with different grammar and words" while *lahjeh* was defined in terms of "pronunciation". Additionally, participants uniformly stated that "whereas understanding a *lahjeh* which belongs to my mother tongue is not challenging, *guyesh* may be unintelligible due to differences at structural and terminology levels in addition to pronunciation". Half of the participants further added that in such cases, the extent of mutual intelligibility depends on "how much of their words I am familiar with". Although Chomsky (1972) argues for detached domain of linguistics from actual speech and the

social setting in which it happens, regardless of their linguistics knowledge, the participants believed that “*zabān*, *guyesh* and *lahjeh* are part of the communities where they are used”.

3.2.4.5. Writing systems and extent of shared words

In their approach to differentiating different types of language varieties, the participants recognized a writing system and the extent to which two varieties share their words as impacting their perspectives about language varieties. On the significance of vocabulary as a fundamental parameter that differentiates types of language varieties, Wolfram and Schilling (2016) assert that “vocabulary is one of the most transparent ways in which dialects differ, and vocabulary studies are a common way in which dialect differences are profiled” (p. 22).

The participants defined *zabān* as “varieties which do not share many of their words and are not mutually intelligible”. They differentiated *zabān* from *guyesh* and *lahjeh* which “share more of their words”. Although the participants were not certain as to what extent of shared words exist between two *guyesh* or two *lahjeh* of the same *zabān*, they confirmed that “two *lahjeh* share almost all of their words”.

Not only the extent of shared words, but whether a language variety has a writing system influenced the participants’ perspectives about language varieties. All of the participants stated that “every *zabān* must have a writing system”. However, they disputed over whether a *guyesh* has a writing system. Abade’i and Tehrani participants argued that “a *guyesh* which is only mutually intelligible to lower extents has a writing system but a *guyesh* which is moderately intelligible does not have a writing system”. Further probing into these perspectives revealed that *guyesh* is not conclusively with or without a writing system. That is, even when a language variety was

determined as a *guyesh* based on other parameters, the participants underlined the extent to which they could understand the speakers of this language variety as an influential parameter to determine whether or not it should have a writing system. Thus, the existence of a standardized writing system, one which is retrievable by everyone who knows this language and without the intervention of the utterer (see Daniels & Bright, 1996, p. 3), for *guyesh*, based on the participants' perspectives, was controversial and dependent on a non-structural parameter.

The participants' idea of a separate system of writing for *guyesh* that are mutually intelligible to lower extents may underline the sociolinguistics of a writing system. That is, writing systems indicate people's tendency to "create a common identity, reject an identity, or construct boundaries and divisions" (Sebba, 2009, p. 39). As Sebba (2009, p. 42) explains, creating distinctiveness "is important particularly for unstandardized language varieties and dialects that need to establish what they are not".

3.2.4.6. Exposure to language varieties

Among parameters underlined as significant, the participants determined exposure to language varieties other than their own as influential in how they define types of language varieties. Regarding this, the participants explained that "exposure to other language varieties leads to familiarity with their vocabulary and grammar". This, in turn, "reinforces higher extents of mutual intelligibility". On the other hand, the participants believed that fewer exposure occasions to language varieties will likely end in lower extents of mutual intelligibility between the speakers of such language varieties. In such situations, the participants differentiated these varieties as "*guyesh* or even *zabān*".

3.2.4.7. Setting and purpose of communication

One of the challenges that the participants encountered in identifying language varieties was due to the imbalanced influence of linguistic and non-linguistic parameters on how they perceived and differentiated different types of language varieties. Some participants believed that “the context in which and the purpose for which a language variety is used” influence how one defines it. Regarding this, Saville-Troike (2003) explains that “the physical setting of an event may call for the use of a different variety of language even when the same general purpose is being served, and when the same participants are involved” (p. 63). Under such circumstances, people’s voice level, non-verbal behaviour and choice of words and structures are different (Saville-Troike, 2003). The participants further explained that a language variety which is used “in the government and/or for political affairs of a country” is the official *zabān* of the respective country. In such cases, they added that “no matter the extent of shared words and/or grammar, this variety is a *zabān*” and mutual intelligibility “has little to no applicability in these situations”.

3.2.4.8. Demographic information

Some of the participants explained that people’s age and education affect how they define types of language varieties. They asserted that “younger well-educated people who know multiple language varieties can define *zabān* and *guyesh*” whereas “illiterate people who have been exposed to their mother tongue and lived in one city their entire life cannot define types of language varieties”. In this regard, Saville-Troike (2003) asserts that description and analysis of varieties of language are associated with social factors such as setting, age, gender, social class, education, and occupation. While the participants’ views concerning the above-mentioned parameters were not

analogous, the quantitative findings confirmed the meaningful contribution of some of these parameters in identifying language varieties in Iran (see Step eleven).

3.2.4.9. Relationships

The participants called attention to social relationships between speakers as a parameter influencing the way they define and use a language variety. They explained that after a certain age, choice of language variety depends on context and relationships. Such responses were in line with Fishman's (1971) concept of domain where one's choice of a variety from their communicative repertoire very much depends on "...topic of communication, relationships between communicators, and locales of communication, in accord with the institutions of a society and the spheres of activity of a speech community" (p. 587). By way of example, the participants established "professor-student relationship at school" and "employer-employee rapport at workplace" as two significant occasions which determine how they define and use a language variety. In these situations, the participants mentioned that they pay attention to their "pronunciation" and "the words" that they use in their conversations. They continued to say that the conversations which take place in formal situations require "careful choice of words and structures". Saville-Troike (2003) also confirms that the roles which people assume in their conversations with others are evident in their choice of pronouns, use of particle, voice pitch or volume, address terms, hesitancy, hedging, and non-verbal behaviour (p. 75).

Kurdish and Abade'i participants further added that whereas they used Farsi "in the education system with the teachers" and "at work with the manager and colleagues", they used their own language variety "with friends, classmates and sometimes even with colleagues".

Saville-Troike (2003, p. 76) further emphasizes that in any given role relationship, the speakers change the features of their speech because “particular role-relationships require that clearly distinct varieties of language be used”.

3.2.4.10. Diglossia and bilingualism

Many of the participants recognized an element of ‘need’ as a significant parameter in how they define and use language varieties specially in bilingual speech communities. According to these participants, bilingual children switch to the official *zabān* of the country (i.e., Farsi) when they see an urgency to do so. For example, the majority of participants referred to schools as “educational settings where teachers, staff and peers may neither understand nor communicate in the language variety with which children are raised”. Under such circumstances, most Kurdish and some Abade’i participants explained that while they use their own variety in “personal matters such as praying, self-talk, and expressing emotions”, they use Farsi for the sole purpose of “communication ‘outside’ the household”. This situation indicates a diglossia where two or more varieties of the same language or two or more different languages within the same speech community are allocated to different functions and contexts (Saville-Troike, 2003, pp. 45-47).

The participants further explained that they see the official *zabān* of the country as a variety with “more acceptability and use” due to “limited applicability” of their own variety “at workplace, educational system and government sectors”. This perspective may mirror sociopolitical priorities of Iranian societies in certain settings such as the government (e.g., see Hassanpour, 1992, 2000). Further, in some speech communities, these perspectives may result in monolingualism among

children and greater society, since children may be more inspired to reject their own language variety (Kulick, 1992 as cited in Ochs & Schieffelin, 2006).

The participants further added that “some families tend to use Farsi as the first language variety when communicating with children”. Though most Kurdish participants considered speaking one’s own language variety as a priority at home, some Abade’i and Tehrani participants admitted that “in pursuit of great educational opportunities or for job security, speaking Farsi with children should be encouraged at home”. In their study, Anonby and Yousefian (2011) publish similar results with popularity and centrality of certain language varieties among families at home since life revolves around these varieties.

3.2.4.11. Language and identity

The majority of participants identified their language variety “as marker of identity attached to speech community”. Language is not only a tool for communication. It is the means speakers use to claim nationhood, a tool with which the strangers are singled out and speech community members affiliate themselves with (Joseph, 2004; Rezaei, Khatib, et al., 2014). Kurdish participants explained that fluency in Farsi was not enough reason to deem it as the language variety to which they had a sense of belonging. According to these participants, “language skills such as reading and writing Farsi has little to do with identifying with it”.

A few of Abadei’ and Tehrani participants, on the other hand, claimed that linguistic identity could be affected by “exposure to a second and third language variety”. These participants underlined their sense of belonging to “more than just one language variety”, since linguistic identity is an ‘ever-changing package’ co-constructed in relation to other people (Odango, 2015).

Further, Abadei' participants explained that in certain situations such as “formal meetings at work or meeting new people in social gatherings” they identify with Tehrani Farsi as the prestigious language variety of the country. Windfuhr (1995) underlines this as an act which speakers of socially lower status tend to do.

3.2.4.12. Religious affiliation

All participants identified their “religious priorities” as affecting their linguistic perspectives. Abade'i participants pointed to speech communities such as their own where “individuals follow religious denominations challenged by the majority of community members”. In such societies, “to reinforce boundaries and maintain the unity of religious community”, minority speech community members may use their own language variety “in different settings and for different purposes”. Some of the community members, however, refrain from using their variety “to avoid recognition with such religious beliefs” because, as scholars (e.g., see Hemming & Madge, 2011; Souza, 2016; Yaeger-Dror, 2014) state, speakers' linguistic choices reflect not only their affiliation to their religious community of practice, but also the strength of their ideological commitment to a specific religion.

In summary, the qualitative exploration of folk linguistic perspectives determined multiple parameters as meaningful in language identification. These results underlined the varied perspectives of Kurdish, Tehrani and Abade'i participants of this study regarding the relevance of geography (urban vs. rural) and number of speakers in language identification. On the other hand, the majority of interviewees reached an agreement regarding the significance of the status of language varieties, the extent of mutual intelligibility and shared words, writing system, exposure

to language varieties, context and purpose of communication, one's demographic information, social relationships, and language and religious identity in defining and differentiating different types of language varieties.

4. Hybrid thesis: Part two (Phase two)

After the assembly of an expert panel (Step one) and the qualitative exploration of folk linguistic perspectives (Step two), with the purpose of validating the questionnaire, I initiated the second phase. This phase was completed in eight steps. In **Step three**, based on the available themes and multiple sessions with the expert panel, I generated seven components and 36 items. In **Step four**, I determined the scaling of the statements. I utilized a 5-point scale form of Likert scale as the most popular one. In **Step five**, I developed the preliminary demographic section of the questionnaire. In **Step six**, to investigate internal consistency, face and content validity, the expert panel checked all the items. Due to sociopolitical priorities practiced in Iran, redundancy, ambiguity, length, and limited relevance, the number of items was substantially reduced to 30 items presented in six components. In **Step seven**, to investigate how interview and questionnaire data integrated, I ran the first pilot. In **Step eight**, analysis and integration of qualitative and quantitative data via methodological triangulation revealed significant results regarding definition and differentiation of different types of language varieties with respect to folk perspectives. Following the developmental procedure of a measurement tool, to ensure that the questionnaire generated consistent responses and was reliable, in **Step nine**, I ran the second and third pilot iterations. The results from these two pilots were, for the most part, consistent. Having established the consistency of results via pilot iterations, in **Step ten**, I initiated the validation process. Statistical estimates indicated that there were no concerns regarding the validity of the questionnaire. Results from phase two were presented in a paper entitled “Language identification

in Iran: A sociolinguistic model and questionnaire”. This paper was accepted (with revisions) in the journal *Language Related Research (LRR)*.

4.1. Phase two- Step three: Components of the questionnaire and item generation

As researchers assert (e.g., see Hymes, 1964, 1974, Saville-Troike, 2003), the value and priority assigned to influential parameters in defining and differentiating types of language varieties are not necessarily similar within, not to mention, across speech communities. With the complexities around validating a questionnaire in mind and considering the multitude of intertwined parameters at work, I included components raised based on the themes identified by the participants, reviewed in the available literature, and agreed upon by the expert panel. As Hinkin (1998) asserts, it is never possible to investigate the complete domain of interest, but it is important that the pool of generated items represent the construct under examination. To address the issues of language identification with respect to folk perspectives and in light of the realities of Iran as a multilingual country, I focused on seven components (C1 to C7) that recurred in the interview process (see Table 19). Due to sociopolitical priorities of Iran as a multilingual and multireligious country (Moradi, 2019), the panel of experts deemed themes related to people's linguistic and religious identities sensitive and advised against developing them into components.

Table 19. Components of the questionnaire based on qualitative themes

Component	Definition
C1-GE. Geography	This component reflects participants' perspectives regarding whether urban vs. rural geographical territories where language varieties are spoken are influential in language identification.
C2-NS. Number of Speakers	This component reflects participants' perspectives regarding whether the number of people who speak language varieties are influential in language identification.
C3-ST. Status	This component indicates whether the status of a language variety in education system and government influences participants' linguistic perspectives.
C4-MI. Mutual Intelligibility	This component shows the extent to which mutually understanding language varieties helps participants define and differentiate the types of these language varieties.
C5-SW. Shared Words	This component shows the extent to which sharing words with other language varieties influences participants' linguistic perspectives.

C6-WS. Writing System	This component focuses on whether having a writing system influences participants' perspectives about language varieties.
C7-EX. Exposure	This shows how exposure to language varieties other than one's own variety will influence participants' perspectives about the respective varieties.

4.1.1. C1-GE (Geography)

The statements included in component one invited participants to indicate whether geography (urban vs. rural) was relevant in language identification. During the qualitative phase, the participants had varied perspectives regarding this parameter as a potential influence on how people define and differentiate different types of language varieties. The participants took varied approaches about the significance of these parameters in the hierarchy of non-linguistic parameters. After consulting with the expert panel, one of the identified components in the questionnaire was derived from this parameter.

4.1.2. C2-NS (Number of Speakers)

Similar to their perspectives about the relevance of geography in defining and differentiating the types of language varieties, the participants expressed heterogeneous perspectives regarding whether the number of speakers meaningfully contributed to language identification. Although the prominence of this parameter in identifying language varieties was not agreed upon by all the interviewees, the second component (C2-NS) explored pilot populations' perspectives about the statements related to the significance of number of speakers in defining and differentiating different types of language varieties.

4.1.3. C3-ST (Status)

In the third component (C3-ST), statements related to the status of language varieties according to the government and in the education system. As scholars assert, status of a language variety according to the government (e.g., see Hassanpour, 1991, 1993, 2000), its use in the education system (e.g., see Adger et al., 1993; Adger & Christian, 2007; Elling, 2013) and social media (e.g., see Moradi, 2019; Rezaei et al., 2017; Saleh, 2013) combine to create a controversial and complicated issue. These challenges more likely add to the complexities of language identification. Therefore, I included statements which examined whether language varieties including *guyesh* and *lahjeh* have an official status according to the Iranian government and in the education system.

4.1.4. C4-MI (Mutual Intelligibility)

Mutual intelligibility – the degree to which speakers understand one another – was a further important parameter and was treated in the questionnaire’s fourth component (C4-MI). This parameter has been used by scholars as well as non-specialists as a central criterion in language identification (Eberhard et al., 2021; Modarresi, 2014). Thus, the statements included in this component investigated how people identify different types of language varieties based on this parameter.

4.1.5. C5-SW (Shared Words)

Mutual intelligibility is also correlated, albeit imperfectly, to the proportion of structures shared between languages. While phonological and grammatical variation also affect intelligibility to some extent, shared lexicon (vocabulary) is foundational to it, and speakers have a strong

explicit awareness of variation in vocabulary (Gooskens & Schneider, 2016; Gooskens et al., 2018). In this questionnaire, component five (C5-SW) therefore investigated how sharing words with other varieties impacts people's identification of language varieties. The decision in establishing the extent of shared words was informed by interviewees' own comments on what proportions of words two *zabān*, two *guyesh*, or two *lahjeh* shared. In other words, to understand the relevance of the extent of shared words for speakers, I did not refer to a technical definition of cognate percentages based on standardized wordlists, or introduce finer distinctions such as different degrees of similarity for different domains of vocabulary such as core vs. peripheral or inherited vs. borrowed words. The interviewees believed that if they could understand only half of the words, without distinguishing what types of words, when having everyday conversations with speakers of another language variety, the respective variety is a *zabān*.

Developing *guyesh*- and *lahjeh*-related statements was inspired by Eberhard's et al. (2021) views with respect to lexical similarity in *Ethnologue*. *Ethnologue* establishes the percentage of lexical similarity between language varieties by estimating the number of similar forms based on standardized wordlists: "Percentages higher than 85% usually indicate a speech variant that is likely a dialect of the language with which it is being compared". Although I did not use wordlists in calculating such percentages, I used a similar percentage (80%) to determine whether folk identify language varieties which share more than 80% of their words as two *guyesh* or two *lahjeh*. To develop equivalent linguistic labels, such estimates can help researchers find out which of these Farsi terms better correspond to the English language idea of 'dialect'.

4.1.6. C6-WS (Writing System)

Component six examined the relevance of a standardized writing system in the definition and differentiation of different types of language varieties. The statements included in this component of the questionnaire investigated whether the existence of a standardized writing system is prerequisite, or sufficient, to consider a variety a *zabān*. Further, this component examined whether similar to the perspectives of “the gatekeepers of written tradition” (see Fishman, 1977, p. XVI; Geerts et al., 1977, p. 206), only the standard dominant languages can have prescriptive and authoritative orthographies and writing systems (see Jaffe, 2000).

4.1.7. C7- EX (Exposure)

People’s perspectives about language varieties are directly affected by exposure to the respective varieties (Chambers & Trudgill, 1998; Miller et al., 2006). During the qualitative phase, the participants identified exposure to language varieties other than their own as a parameter which helped expand the extents of mutually understanding these varieties. This influenced their perspectives about different types of language varieties. As such, statements in component seven of this questionnaire investigated whether exposure to language varieties had an impact on how they defined and differentiated different types of language varieties.

Having established the components, the next step was to verbalize the concepts to be measured and generate items which participants could use to express, as closely as possible, their views about the phenomenon under study. As such, qualitative data was transformed into and led to the generation of a pool of 36 items which could be quantitatively assessed during the pilot iterations. It is noteworthy that among the concepts to examine was folk attitude towards language

variety types including *zabān*, *guyesh* and *lahjeh*. In other words, I originally aimed to develop a separate section which could investigate language varieties that participants would consider pleasant, correct and useful, compared with other variety types. However, considering limits (e.g., length of the questionnaire) and challenges (e.g., validating a measurement instrument), not all of the items which could examine folk perspectives about pleasantness, correctness and usefulness of language varieties were included in the pool of items.

To avoid linguistically loaded items and ensure simplicity and ease of comprehension, I consulted the expert panel. From among the presented items, the panel agreed upon 30 items for the first pilot iteration (see Step six). All the items were finalized based on themes and experts' views in consultation with the available literature on language identification and parameters iterated as meaningfully significant. I understood the depiction of each of the items on the components. I disclosed the representation of each item on the components (see Table 20).

Table 20. Pool of generated items based on the components

Component	Item
C1-GE. Geography	1 People in villages generally speak using <i>guyesh</i> (dialects).
	2 People in villages generally speak with <i>lahjeh</i> (accents).
	3 People in cities generally speak <i>zabān</i> (languages).
	4 People who speak a variety of my <i>zabān</i> (language) but live in a different geographical region, probably speak a different <i>guyesh</i> (dialect).
C2-NS. Number of Speakers	5 A <i>guyesh</i> (dialect) is always part of a larger <i>zabān</i> (language) because it has fewer speakers.
	6 If a language variety has only a few thousand speakers, it should generally be considered a <i>guyesh</i> (dialect).
	7 If a language variety has at least a million speakers, it should generally be considered a <i>zabān</i> (language).
C3-ST. Status	8 <i>Guyesh</i> (dialect) is a language variety which, according to the government, has no official status.
	9 <i>Guyesh</i> (dialect) is a language variety that is not used in the education system.
	10 <i>Lahjeh</i> (accent) is a language variety that is not used in the education system.
	11 <i>Lahjeh</i> (accent) is a language variety which, according to the government, has no

	<p>official status.</p> <p>12 There should be a linguistic label between <i>zabān</i> (language) and <i>guyesh</i> (dialect) to account for all the language varieties that are neither of these two language varieties.</p> <p>13 Farsi is the most useful language variety.</p> <p>14 Farsi is the most prestigious language variety.</p> <p>15 Farsi is the language variety used in formal settings such as the government.</p> <p>16 I use my own language variety to speak with family and friends.</p> <p>17 To be considered <i>zabān</i> (language), a language variety must be the official <i>zabān</i> (language) of some country.</p>
C4-MI. Mutual Intelligibility	<p>18 If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate <i>zabān</i> (languages).</p> <p>19 I cannot understand a language variety if I am not familiar with its words.</p> <p>20 If someone speaks to me in a <i>lahjeh</i> (accent) of my mother tongue which is different from my own <i>lahjeh</i> (accent), it is always possible to understand them quite well.</p> <p>21 If someone speaks to me in a <i>guyesh</i> (dialect) of my mother tongue which is different from my own <i>guyesh</i> (dialect), it is always possible to understand them quite well.</p> <p>22 Partial intelligibility between speakers of two language varieties means that these varieties belong to the same <i>zabān</i> (language).</p> <p>23 High extents of mutual intelligibility between speakers of two language varieties make these varieties <i>guyesh</i> (dialect) of the same <i>zabān</i> (language).</p> <p>24 Low extents of mutual intelligibility between speakers of two language varieties make these varieties separate <i>zabān</i> (languages).</p>
C5-SW. Shared Words	<p>25 If two language varieties share less than 50% of their words, they are generally separate <i>zabān</i> (languages).</p> <p>26 If two language varieties share more than 80% of their words, they are generally <i>lahjeh</i> (accents) of the same <i>zabān</i> (language).</p> <p>27 If two language varieties share more than 80% of their words, they are generally <i>guyesh</i> (dialects) of the same <i>zabān</i> (language).</p> <p>28 To decide whether two language varieties are <i>zabān</i> (languages), <i>guyesh</i> (dialects) or <i>lahjeh</i> (accents), the most important parameter is the extent of words that they share.</p>
C6-WS. Writing System	<p>29 If a language variety has a standardized writing system, it may not be a <i>zabān</i> (language).</p> <p>30 <i>Guyesh</i> (dialects) do not have standardized writing systems.</p> <p>31 To be considered a <i>zabān</i> (language), a language variety needs to have a standardized writing system.</p> <p>32 <i>Guyesh</i> (dialects) which are mutually intelligible to lower extents have standardized writing systems.</p> <p>33 <i>Guyesh</i> (dialects) which are mutually intelligible to some extents do not have standardized writing systems.</p>
C7-EX. Exposure	<p>34 The more I am exposed to a new language variety, the better I understand it.</p> <p>35 All the people who live in Iran are exposed to Farsi.</p> <p>36 I need to be exposed to a language variety to decide whether it is a <i>zabān</i> (language), <i>guyesh</i> (dialect) or <i>lahjeh</i> (accent).</p>

4.2. Phase two- Step four: Rating scales of the items

Having generated the items, the next step was to determine the scaling of the statements. The rating scale utilized in this study was based on Likert scale. After reviewing the available literature on questionnaire development (e.g., see Hinkin, 1998; Neuman & Robson, 2012) and a session with the expert panel, I realized that it is a common practice to employ six-, seven- and nine-point scales. However, Lissitz and Green (1975) argue that coefficient alpha reliability levels off after five-point scales.

Further, considering the diversity of the language situation across the country and the topical specificity of some of the questions, I hypothesized that the ‘don’t know’ or ‘undecided’ option would be an important element. Despite the mixed literature on providing a neutral category (e.g., see Allen & Seaman, 2007; Armstrong, 1987; Dörnyei & Taguchi, 2010), I established a rating scale which granted participants uncertainty of choice. The expert panel also confirmed that it was important to include a midpoint because a substantial minority of people hold attitudinal positions which are genuinely neutral. Omitting a midpoint would result in such individuals being forced to select an alternative response that does not reflect their true position on the underlying preference dimension. In choosing the number of categories, panel members underlined the possibility to combine or collapse data even after data collection in case specification of data was needed. Thus, I opted for a five-option type rating scale: agree, somewhat agree, somewhat disagree, disagree, and don’t know.

4.3. Phase two- Step five: Developing the preliminary demographic questions

In addressing the issues of language identification, demographics of the participants have been abundantly emphasized as potential influential parameters (e.g., see Preston, 1989; Saville-Troike, 2003). Hence, they were included in a separate demographics section. External variables featured in the preliminary section of the questionnaire included participants' mother tongue, additional languages, education, formal linguistic knowledge, gender, place of birth, and age. Such information allowed for analysis of variation and predictive influence of external variables in the way that people identified language varieties (see Step eleven).

4.4. Phase two- Step six: Item checking via expert review

The questionnaire with seven components and 36 items was given to an expert panel. Among the goals of this panel was an analysis of this questionnaire to accommodate folk perspectives about different types of language varieties. They discussed the validity of the questionnaire as a tool for systematic representation and organization of language identification data. Not only did they evaluate it as a whole, but they also examined the intelligibility of individual items for potential participants without an academic background as well as educated participants without a background in linguistics.

Considering that the questionnaire focused on multiple concepts (see Table 20), the experts felt that it might be unintelligible in some areas, even though most concepts were expressed in non-technical language. Due to redundancy and ambiguity, length and limited relevance, the number of items was reduced from 36 to 30, limited to those directly significant to addressing language identification issue.

Additionally, equivalency of Farsi and English terms was reviewed by these experts, three of whom work in both languages. As the literature suggests, definitions of linguistic labels have not always been agreed upon (e.g., see Dabir-Moghaddam, 2008; Modarresi, 2014; Sadeghi, 1970). While linguists draw on their linguistics expertise in this and provide clear-cut definitions at times, based on the criteria outlined in the conceptual framework section above, sociolinguists and folk linguists reiterate the prominence of all people's linguistic perspectives. As a result, providing fully equivalent linguistic labels in translation, and identifying firm borderlines between different types of language varieties, was not easily achievable.

Considering the ongoing challenges with regard to definition and differentiation of language variety types, the expert panel faced a decision-making problem about providing equivalent linguistic labels for terms such as *zabān*, *guyesh* and *lahjeh*. Regarding this, panel members explained that distinguishing *guyesh* and *lahjeh* alone may be challenging enough for people without linguistics knowledge as these are linguistic jargon generated by linguists (Dabir-Moghaddam, 2008). Regarding this issue, the pilot iteration results addressed the panel's concern. Results revealed that the participants were able to differentiate among the language variety types (see Step nine). The experts further emphasized creating statements which avoid feeding the participants new linguistics-related knowledge. The significance of this part was particularly noticeable in certain statements where the participants were to differentiate linguistic terms such as *zabān*, *guyesh* and *lahjeh*.

Not only equivalency of English-Farsi linguistic terms, but also the overall translation of the questionnaire was reviewed by the panel members. To ensure accuracy of translation, I

translated the model following forward and backward approaches to translation (i.e., back translation). As such, the questionnaire was translated from and to the target language (Farsi). This process enabled me to control for misunderstandings and confusion caused by wording. Having completed the translation process, the pre-final version of the questionnaire was reviewed by the experts. These experts reviewed the translations, achieved consensus on equivalency and resolved discrepancies.

Although, to a major extent, reliant on the themes generated from the qualitative phase of this study, in the past few years, the items were refined thanks to the panel's expertise. Such unanimous refinements took place to be mindful of sensitive sociopolitical priorities among speech community members. The experts' systematic analysis led to a questionnaire with 30 items and six components, namely: geography, number of speakers, status, mutual intelligibility, extent of shared words, and writing system (see Table 21). The component 'exposure' was deleted due to sociopolitical sensitivities involved with this component and representative items. It is noteworthy that while I consulted the panel for guidance in all three phases of this study, ultimately it was the validation of the questionnaire during Step ten as well as research results that confirmed which of the design decisions and hypotheses were borne out.

Table 21. Refined questionnaire based on experts' views

Component	Item
C1-GE. Geography	1 People in villages generally speak using <i>guyesh</i> (dialects).
	2 People in villages generally speak with <i>lahjeh</i> (accents).
	3 People in cities generally speak <i>zabān</i> (languages).
	4 People who speak a variety of my <i>zabān</i> (language) but live in a different geographical region, probably speak a different <i>guyesh</i> (dialect).
C2-NS. Number of Speakers	5 A <i>guyesh</i> (dialect) is always part of a larger <i>zabān</i> (language) because it has fewer speakers.
	6 If a language variety has only a few thousand speakers, it should generally be considered a <i>guyesh</i> (dialect).
	7 If a language variety has at least a million speakers, it should generally be considered a <i>zabān</i> (language).
C3-ST. Status	8 <i>Guyesh</i> (dialect) is a language variety which, according to the government, has no official status.
	9 <i>Guyesh</i> (dialect) is a language variety that is not used in the education system.
	10 <i>Lahjeh</i> (accent) is a language variety that is not used in the education system.
	11 <i>Lahjeh</i> (accent) is a language variety which, according to the government, has no official status.
	12 There should be a linguistic label between <i>zabān</i> (language) and <i>guyesh</i> (dialect) to account for all the language varieties that are neither of these two language varieties.
	13 Farsi is the most useful language variety.
	14 To be considered <i>zabān</i> (language), a language variety must be the official <i>zabān</i> (language) of some country.
C4-MI. Mutual Intelligibility	15 If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate <i>zabān</i> (languages).
	16 I cannot understand a language variety if I am not familiar with its words.
	17 If someone speaks to me in a <i>lahjeh</i> (accent) of my mother tongue which is different from my own <i>lahjeh</i> (accent), it is always possible to understand them quite well.
	18 If someone speaks to me in a <i>guyesh</i> (dialect) of my mother tongue which is different from my own <i>guyesh</i> (dialect), it is always possible to understand them quite well.
	19 Partial intelligibility between speakers of two language varieties means that these varieties belong to the same <i>zabān</i> (language).
	20 High extents of mutual intelligibility between speakers of two language varieties make these varieties <i>guyesh</i> (dialect) of the same <i>zabān</i> (language).
	21 Low extents of mutual intelligibility between speakers of two language varieties make these varieties separate <i>zabān</i> (languages).
C5-SW. Shared Words	22 If two language varieties share less than 50% of their words, they are generally separate <i>zabān</i> (languages).
	23 If two language varieties share more than 80% of their words, they are generally <i>lahjeh</i> (accents) of the same <i>zabān</i> (language).
	24 If two language varieties share more than 80% of their words, they are generally <i>guyesh</i> (dialects) of the same <i>zabān</i> (language).
	25 To decide whether two language varieties are <i>zabān</i> (languages), <i>guyesh</i> (dialects) or <i>lahjeh</i> (accents), the most important parameter is the extent of

	words that they share.
C6-WS. Writing System	<p>26 If a language variety has a standardized writing system, it may not be a <i>zabān</i> (language).</p> <p>27 <i>Guyesh</i> (dialects) do not have standardized writing systems.</p> <p>28 To be considered a <i>zabān</i> (language), a language variety needs to have a standardized writing system.</p> <p>29 <i>Guyesh</i> (dialects) which are mutually intelligible to lower extents have standardized writing systems.</p> <p>30 <i>Guyesh</i> (dialects) which are mutually intelligible to some extents do not have standardized writing systems.</p>

Moreover, the number of items included in each of the respective components varied between three to seven. Although there were no strict rules regarding this, I was content with the generated items since the aim was to create a short measure “as an effective means of minimizing response biases caused by boredom or fatigue” (Hinkin, 1998, p. 109). Furthermore, scholars (e.g., see Cortina, 1993; Hinkin, 1998) point out that in developing scales with internal consistency, as few as three to four items per component suffice to analyze the homogeneity of items with each latent construct. As Hinkin (1998) asserts, final determination can only be achieved via evidence of construct validity for measures (see Step ten).

4.5. Phase two- Step seven: First pilot

A challenging step in developing an independently filled-out questionnaire is conceptualization of constructs. Conceptualization of constructs such as human thinking or attitude is a complex process which requires researchers to clarify in their research what they mean explicitly for others to see (Neuman & Robson, 2012). In this process, I relied on multiple sources including personal experience and available literature. In the case of this project, I collected qualitative data prior to pilot iterations and consulted available literature as well as the expert panel. Not only is a clear definition of constructs required, but the generation of statements which

best exemplify the constructs in non-technical language and understandable to all participants from various communities, is also essential to create valid findings. Considering that the questionnaire items were finalized based on the panel’s perspectives and expertise, I was curious as to how interview and questionnaire data integrated and yielded results regarding definition and differentiation of language variety types from people’s perspectives. For this reason, the first pilot was run. It is noteworthy that Section 4.5.1. (geography) to Section 4.5.6. (writing system) present first pilot data analysis results. Significance and interpretation of these results as well as how they integrate with the qualitative results are discussed in Section 4.6. (Step eight). Also, as mentioned earlier (see Section 2.3.2.), ‘partly agree’ and ‘agree’ as well as ‘partly disagree’ and ‘disagree’ are combined in data analysis.

4.5.1. C1-GE (Geography)

As Table 22 shows, via four geography-related statements, C1-GE sought participants’ perspectives on the relevance of geography (urban vs. rural) in language identification. Results showed that an average of 44% of the participants disagreed with C1.1-C1.4, an average of 30% of the participants agreed with them and an average of 22% chose ‘don’t know’ in this regard.

Table 22. C1-GE (Geography): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C1.1. People in villages generally speak using guyesh (dialects).	33.3%	13.3%	20.0%	16.7%	16.7%
C1.2. People in villages generally speak with lahjeh (accents).	26.7%	6.7%	23.3%	26.7%	16.7%
C1.3. People in cities generally speak zabān (languages).	13.3%	10.0%	30.0%	46.7%	0.0%
C1.4. People who speak a variety of my zabān (language) but live in a different geographical region, probably speak a different guyesh (dialect).	6.7%	10.0%	16.7%	13.3%	53.3%

Nearly 47% of the participants agreed that people in villages generally speak using *guyesh* (C1.1) whereas 33.4% of them agreed that in villages people speak with *lahjeh* (C1.2). The majority of participants (77%) disagreed with C1.3, that people in cities generally speak *zabān*. More than half of the participants (53.3%) chose ‘don’t know’ regarding C1.4, that people who speak a variety of one’s own *zabān* but live in a different geographical region probably speak a different *guyesh* (see Table 22).

4.5.2. C2-NS (Number of Speakers)

With respect to the significance of the number of people who speak different types of language varieties in language identification, an average of 49% of participants disagreed and 39% agreed with the three statements in this component. To further elaborate, nearly 47% of these participants disagreed with C2.1, that a *guyesh* is always a part of a larger *zabān* because it has fewer speakers. Almost 47% of the participants agreed with C2.2, that a language variety with only a few thousand speakers should generally be considered a *guyesh*. Lastly, 57% disagreed with C2.3, that a language variety with at least a million speakers should generally be considered a *zabān* (see Table 23).

Table 23. C2-NS (Number of Speakers): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C2.1. A <i>guyesh</i> (dialect) is always part of a larger <i>zabān</i> (language) because it has fewer speakers.	10.0%	26.7%	6.7%	40.0%	16.7%
C2.2. If a language variety has only a few thousand speakers, it should generally be considered a <i>guyesh</i> (dialect).	16.7%	30.0%	20.0%	23.3%	10.0%
C2.3. If a language variety has at least a million speakers, it should generally be considered a <i>zabān</i> (language).	13.3%	20.0%	30.0%	26.7%	10.0%

4.5.3. C3-ST (Status)

C3-ST statements (see Table 24) investigated the status of language variety types within the educational system and according to the government. Findings of the first pilot in this section revealed that 77% and 47% of the participants respectively agreed that, according to the government, *guyesh* and *lahjeh* have no official status. With respect to use in the education system, while the majority of participants disagreed that *lahjeh* is not used in the education system (C3.3), only one-third of them held the same view about *guyesh* (C3.2). Nearly half of the participants chose ‘don’t know’ about C3.5, that there should be a linguistic label between *zabān* and *guyesh* to account for all the language varieties that are neither of these two language varieties. Lastly, two-thirds of the sample population agreed with C3.6, that Farsi is the most useful language variety) and C3.7, that to be considered *zabān*, a variety must be the official *zabān* of some country (see Table 24).

Table 24. C3-ST (Status): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C3.1. <i>Guyesh</i> (dialect) is a language variety which, according to the government, has no official status.	43.3%	33.3%	0.0%	3.3%	20.0%
C3.2. <i>Guyesh</i> (dialect) is a language variety that is not used in the education system.	13.3%	23.3%	13.3%	23.3%	26.7%
C3.3. <i>Lahjeh</i> (accent) is a language variety that is not used in the education system.	13.3%	10.0%	50.0%	26.7%	0.0%
C3.4. <i>Lahjeh</i> (accent) is a language variety which, according to the government, has no official status.	10.0%	36.7%	13.3%	10.0%	30.0%
C3.5. There should be a linguistic label between <i>zabān</i> (language) and <i>guyesh</i> (dialect) to account for all the language varieties that are neither of these two language varieties.	6.7%	13.3%	16.7%	16.7%	46.7%
C3.6. Farsi is the most useful language variety.	16.7%	50.0%	26.7%	6.7%	0.0%
C3.7. To be considered <i>zabān</i> (language), a language variety must be the official <i>zabān</i> (language) of some country.	10.0%	6.7%	36.7%	26.7%	20.0%

4.5.4. C4-MI (*Mutual Intelligibility*)

C4-MI statements investigated whether mutual intelligibility among speakers of different language varieties influenced their perspectives regarding these varieties. The majority of participants agreed with C4.1, that if two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate *zabān*. All of the participants agreed with C4.2, that they cannot understand a language variety if they are not familiar with its words. All participants agreed that it is always possible to understand someone who speaks with them in a *lahjeh* of their mother tongue (see C4.3). However, half of them disagreed that if someone speaks to them in a *guyesh* of their mother tongue, it is always possible to understand them (see C4.4). In identifying language varieties based on partial (C4.5) or high (C4.6) extents of mutual intelligibility between speakers of two language varieties, nearly half of the participants chose ‘don’t know’. The majority of participants, however, agreed with C4.7, that low extents of mutual intelligibility between speakers of two language varieties make these varieties separate *zabān* (see Table 25).

Table 25. C4-MI (Mutual Intelligibility): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C4.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate <i>zabān</i> (languages).	33.3%	30.0%	16.7%	6.7%	13.3%
C4.2. I cannot understand a language variety if I am not familiar with its words.	76.7%	23.3%	0.0%	0.0%	0.0%
C4.3. If someone speaks to me in a <i>lahjeh</i> (accent) of my mother tongue which is different from my own <i>lahjeh</i> (accent), it is always possible to understand them quite well.	66.7%	33.3%	0.0%	0.0%	0.0%
C4.4. If someone speaks to me in a <i>guyesh</i> (dialect) of my mother tongue which is different from my own <i>guyesh</i> (dialect), it is always possible to understand them quite well.	6.7%	0.0%	26.7%	26.7%	40.0%
C4.5. Partial intelligibility between speakers of two language varieties means that these varieties belong to the same <i>zabān</i> (language).	6.7%	13.3%	20.0%	6.7%	53.3%
C4.6. High extents of mutual intelligibility between speakers of two language varieties make these varieties <i>guyesh</i> (dialect) of the same <i>zabān</i> (language).	6.7%	0.0%	26.7%	16.7%	50.0%
C4.7. Low extents of mutual intelligibility between speakers of two language varieties make these varieties separate <i>zabān</i> (languages).	26.7%	36.7%	23.3%	13.3%	0.0%

4.5.5. C5-SW (Shared Words)

C5-SW statements (see Table 26) sought to identify whether the extent of shared words among language varieties impacted people’s perspectives in identifying language varieties. Participants were requested to identify varieties which share less than 50% and those which share 80% or more of their words as *zabān*, *guyesh* or *lahjeh*.

Approximately 47% of the participants agreed with C5.1, that two language varieties which share less than 50% of their words are generally separate *zabān*. Almost 67% agreed with C5.2, that two language varieties which share more than 80% of their words are generally *lahjeh* of the same *zabān*. The majority (75%) of participants disagreed with C5.3, that two language varieties which share more than 80% of their words are generally *guyesh* of the same *zabān*. Lastly, nearly

70% agreed with C5.4, that in identifying language varieties one of the most important parameters is the extent of words that they share (see Table 26).

Table 26. C5-SW (Shared Words): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C5.1. If two language varieties share less than 50% of their words, they are generally separate <i>zabān</i> (languages).	13.3%	33.3%	13.3%	6.7%	33.3%
C5.2. If two language varieties share more than 80% of their words, they are generally <i>lahjeh</i> (accents) of the same <i>zabān</i> (language).	40.0%	26.7%	0.0%	0.0%	33.3%
C5.3. If two language varieties share more than 80% of their words, they are generally <i>guyesh</i> (dialects) of the same <i>zabān</i> (language).	0.0%	0.0%	26.7%	46.7%	26.7%
C5.4. To decide whether two language varieties are <i>zabān</i> (languages), <i>guyesh</i> (dialects) or <i>lahjeh</i> (accents), the most important parameter is the extent of words that they share.	26.7%	43.3%	20.0%	0.0%	10.0%

4.5.6. C6-WS (Writing System)

The last component of the questionnaire sought language identification with respect to the writing system (see Table 27). An equal population of 40% of the participants agreed and disagreed with C6.1, that even if a language variety has a standardized writing system, it may not be a *zabān* (see Table 27). Approximately 53% of the participants disagreed whereas one-third agreed with C6.2, that a *guyesh* does not have a writing system. Nearly 73% agreed with C6.3, that to be considered a *zabān*, a variety must have a standardized writing system. Lastly, in C6.4 and C6.5, almost half of the participants chose ‘don’t know’ in establishing a writing system for *guyesh* in terms of the extents of mutual intelligibility (see Section 4.6., Step eight for how quantitative and qualitative results integrate).

Table 27. C6-WS (Writing System): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C6.1. If a language variety has a standardized writing system, it may not be a zabān (language).	20.0%	20.0%	16.7%	23.3%	20.0%
C6.2. Guyesh (dialects) do not have standardized writing systems.	23.3%	10.0%	33.3%	20.0%	13.3%
C6.3. To be considered a zabān (language), a language variety needs to have a standardized writing system.	36.7%	36.7%	23.3%	3.3%	0.0%
C6.4. Guyesh (dialects) which are mutually intelligible to lower extents have standardized writing systems.	10.0%	10.0%	16.7%	16.7%	46.7%
C6.5. Guyesh (dialects) which are mutually intelligible to some extents do not have standardized writing systems.	3.3%	30.0%	10.0%	0.0%	56.7%

4.6. Phase two- Step eight: Integration of data via methodological triangulation

The methodological triangulation (Denzin, 1978), the combination of methodologies in the study of the same phenomenon, is an essential step to providing a more comprehensive picture about a phenomenon under study compared with independently undertaken qualitative or quantitative studies (Barbour, 1998). This type of triangulation contributes to the validity of research results, creates an in-depth understanding of an unexplored or less explored phenomenon (Fielding & Fielding, 1986), and examines whether the results from the two methods confirm and complement one another (Shih, 1998).

Despite the significance of triangulation technique in social sciences, there is little direction about how it must be applied, mainly due to challenges it brings about for researchers (Farmer et al., 2006). Absence of detailed description about how to carry out this type of triangulation has led to reports which are heavy on the findings with little on how the researchers actually got to those findings (Erzerberger & Prein, 1997). Among valid concerns regarding “genuine integration” (Bryman, 2007, p. 8) of data is the lack of formal education in mixed methods research in social

sciences and humanities. As such, very few mixed methods studies have truly triangulated qualitative and quantitative findings.

In light of the collected qualitative and quantitative data and prior to further pilot iterations, I integrated the said data following the ‘between (or across) methods’ (Denzin, 1978) to integration. Despite difficulties in the act of integration, as Morse (1991) suggests, studies which entail a complementary qualitative or quantitative phase can address the complexity of combining the two paradigms. Following Morse’s (1991) suggestion, and embedded in the design of this research project, qualitative method preceded the quantitative one as preliminary inquiry in developing a measurement scale and quantitative analysis of folk linguistic perspectives. Complexity in integration also arises when studies use methodological triangulation to validate a measurement tool. Validating a newly developed measurement tool was not a concern which needed to be addressed at pilot iterations. In other words, the reliability and validity of the language identification questionnaire were treated at later stages (see Section 4.8., Step ten for the validation process).

As Jick (1979) asserts, some research areas are not better geared to a qualitative- or quantitative-oriented method, since both types provide invaluable data that the other method may have been blind towards. Considering that identification of language varieties is heavily reliant on the linguistic and non-linguistic priorities of the specific communities under study (Saville-Troike, 2003), the accuracy of findings from the qualitative phase could be confirmed by collecting data via a questionnaire bearing on the same phenomenon. Since I was the only researcher and observer during the semi-structured interviews in this research project, utilizing methodological

triangulation was intended not only to investigate whether the two methods of data collection agreed with and confirmed one another, but also to refine my understanding of all aspects of language identification in Iran as a less researched and less explored research problem.

I sorted the findings from both qualitative and quantitative phases in separate sheets. To understand the extent of participants' agreement more clearly and considering that 'somewhat agree' and 'somewhat disagree' could not be interpreted without further data, I merged 'somewhat agree' with agree and 'somewhat disagree' with disagree. The two methods together produced largely consistent and convergent results.

One of the themes underlined during the interviews was related to geography (urban vs. rural). Regarding this, the interviewees developed diverse perspectives where nearly half agreed that geography (urban vs. rural) impacted how they defined and differentiated language variety types. The other half, however, disregarded the mentioned parameter as influential in language identification. To determine whether, based on folk perspectives, this parameter meaningfully contributed to language identification, it was developed into the items included in C1-GE. Using the questionnaire yielded similar results which confirmed the heterogeneity of the participants' perspectives in geography-related statements (see Table 22). Similar to geography, both the interviewees and participants expressed mixed views regarding the relevance of the number of speakers in identifying language varieties (C2-NS). Since the results from both methods converged, there is strong evidence to believe on the consistent results of the two methods. These results were later reviewed by the expert panels (see Step nine).

The interviewees further differentiated different types of language varieties based on their status. That is, *zabān* was regarded as a variety used in the education system and with an official status. *Guyesh* and *lahjeh* were generally referred to as without any official status or use in the education. Whether the participants from different speech communities would confirm such perspectives was investigated via C3-ST statements. Points of convergence included the participants' perspectives about the status of *guyesh* as a variety which, according to the government, had no official status. Regarding the use of *guyesh* in the education system, qualitative and quantitative data were not entirely compatible. In other words, whereas the majority of interviewees referred to *guyesh* as a variety not used in the education system, the participants' perspectives about C3.2 (use of *guyesh* in the education system) were distributed almost evenly between Likert-scale points of agreement and disagreement.

I noticed the highest level of inconsistency between the results of the two methods regarding status and use of *lahjeh*. This is to say that, unlike the qualitative data, the majority of participants disagreed with C3.3, that *lahjeh* is not used in the education system. Regarding its status, nearly half of the participants agreed with the interviewees that *lahjeh* does not have an official status according to the government. Additionally, the lack of a firm borderline among language variety types and the urgency to include a linguistic label between *zabān* and *guyesh* was raised by some interviewees. Though this was later confirmed by 20% of the participants, nearly half of the population chose 'don't know'. The second major discrepancy was observed with respect to *zabān*. Except for the Kurdish interviewees, the rest confirmed that a variety must be

the official *zabān* of a country to be considered a *zabān*. Nearly 65% of the participants disagreed with this perspective in C3.7.

The interviewees' perspectives about mutual intelligibility as a prominent parameter in language identification converged with the participants' perspectives in C4-MI. In both methods, data underlined *zabān* as unintelligible varieties. Integrated results further showed that while *lahjeh* were intelligible, *guyesh* were sometimes difficult to understand. Findings revealed that both low and high extents of mutual intelligibility between speakers could be interpreted as two *guyesh*. Though the statements in C4-MI did not contain items which addressed the underlying reasons for folk approach to the extents of mutual intelligibility in language identification, the interviewees added 'religious affiliations' and 'sociocultural priorities' as complementary data which could differentiate language variety types. In addition, the interviewees stated that the extents of mutual intelligibility between speakers of different language varieties could be moderate (partial) in addition to low or high. Building on this idea, item C4.5 of the questionnaire investigated whether partial intelligibility between speakers could be interpreted as varieties of the same *zabān*. More than 50% of the participants chose 'don't know' which brought into attention an area in need of further exploration. As such, a point of discussion in a subsequent panel discussion was about this item. Finally, both data types correlated mutual intelligibility with the extent of familiarity with the words of a language variety.

The quantitative results in C5-SW confirmed the interviewees' approaches towards the extent of shared words in language identification. Findings determined *zabān* and *lahjeh* as varieties which share lower than 50% and higher than 80% of their words, respectively. Regarding

guyesh, however, neither of the two data types defined and differentiated *guyesh* from *zabān* or *lahjeh*. Integration of results underlined *guyesh* as language varieties which share more words compared to *zabān* and fewer words compared to *lahjeh*. However, as the qualitative data confirmed, these borderlines in differentiating the respective varieties were neither clear nor constant.

On the significance of a standardized writing system in language identification, interview results converged with the participants' views in the last component (C6-WS). This is to say that *zabān* was identified as a variety which must have a standardized writing system. *Guyesh*, however, was regarded as a variety which may or may not have a standardized writing system. Although approximately 50% of the participants chose 'don't know' in C6.3 and C6.4, the interviewees had asserted that whether *guyesh* has a standardized writing system relied on the extent to which they were mutually intelligible. As such, folk perspectives about *guyesh* with respect to a writing system were not conclusive.

Literature on language identification indicated that the number, priority and contribution weight of parameters in identifying language varieties in different speech communities are neither comparable nor measurable. As such, whether parameters associated with folk perspectives were parallel was not measurable via either of the methods. Further, the process of compiling research material based on multiple methods was useful whether there was convergence or dissonance. Where there was convergence, as was the case with many of the themes generated during the qualitative phase, confidence in the results grew considerably. In my investigation of the issues of language identification, considering the lack of available literature on language identification or a

quantitative measure to statistically establish the significance of folk perspectives, qualitative data and analysis functioned as the glue that cemented the interpretation of results. As Jick (1979) explains, qualitative data generated ‘holistic work’ or ‘thick description’. Furthermore, triangulation results reassured me that qualitative results do not reflect my subjective biases as a researcher trying to find a solution to the problem at hand.

One of the shortcomings in the triangulation process was related to the association between folk demographics and their linguistic perspectives. Parameters such as education and formal linguistic knowledge seemingly correlated with people’s perspectives in both the qualitative and quantitative phases. However, to ensure whether they statistically significantly contributed to language identification, a measurement tool was required. Though literature has abundantly identified the predictive power of demographic information in folk perspectives about language varieties, at this point of research, integration of qualitative and quantitative results could not determine whether their mother tongue, formal linguistic knowledge, education, age, and gender made a significant contribution.

Points of divergence were addressed via the second and third pilots (see Step nine in Section 4.7.). Though one of the goals of triangulation was to identify points at which data converged, dissonance between data types was not an unexpected or undesirable finding. It confirmed the literature whereby some speech communities may or may not agree on parameters as influential in language identification. In cases where different methods reveal contradictory results, Morse (1991) recommended that the primary method data be the one based on which the researcher makes decisions about the phenomenon under study. Nonetheless, decision making

regarding some of the items, particularly those in the first two components (C1-GE and C2-NS), were delayed until after I ran two more pilot iterations. Before and after further pilots, the integrated results were discussed with the members of the expert panel.

4.7. Phase two- Step nine: Pilots 2 and 3

Social science research has underlined the significance of establishing accuracy and consistency of measurement tools including questionnaires (Hinkin, 1998; Neuman & Robson, 2012; Shultz et al., 2014). However, “dearth of skills and knowledge of validity and reliability test analysis among social and health science researchers” (Bolarinwa, 2015, p. 195) has made any type of test analysis challenging. Moreover, quantification of abstract and intangible concepts such as ‘linguistic perspectives’ and creating a bridge between such concepts and data added to challenges that I had to face. As such, one of my tasks was to determine the reliability and validity of the questionnaire.

As Hinkin (1998) points out, in the process of scale development, multiple small independent samples are beneficial in establishing the content validity. As such, before running Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) with larger samples, to ensure that the pilot questionnaire generated consistent responses and was, therefore, reliable, two significant measures were taken. I addressed the issues (built-in errors) of the questionnaire via multiple panel discussions. Each time and after the issues were resolved, I piloted this questionnaire (a total of three times) to establish that responses were replicable. As a result, I developed a tool that, to the best of my and the experts’ knowledge and specialty, was reliable in measuring folk perspectives about language varieties.

Not only reliability of a measurement tool, but also face and content validity can be established via an expert panel who explore theoretical constructs (Shultz et al., 2014). Thus, the experts investigated whether items included in the questionnaire adequately measured the constructs. Before further pilots and based on an integration of the interview and first pilot's results, the experts examined if the questionnaire at hand was truly representative of a language questionnaire. As a result of multiple sessions with the experts, I decided to further delete a few of the items from the questionnaire that were unanimously confirmed as redundant (i.e., C3.7, C4.5, C4.6, C4.7, C5.4), ambiguous (i.e., C1.4, C6.4, C6.5) and with limited relevance (i.e., C3.6). As a result, the number of items was reduced from 30 to 21 (see Table 28).

Table 28. Language identification questionnaire with 21 items

Component	Item
C1-GE. Geography	1 People in villages generally speak using <i>guyesh</i> (dialects).
	2 People in villages generally speak with <i>lahjeh</i> (accents).
	3 People in cities generally speak <i>zabān</i> (languages).
C2-NS. Number of Speakers	4 A <i>guyesh</i> (dialect) is always part of a larger <i>zabān</i> (language) because it has fewer speakers.
	5 If a language variety has only a few thousand speakers, it should generally be considered a <i>guyesh</i> (dialect).
	6 If a language variety has at least a million speakers, it should generally be considered a <i>zabān</i> (language).
C3-ST. Status	7 <i>Guyesh</i> (dialect) is a language variety which, according to the government, has no official status.
	8 <i>Guyesh</i> (dialect) is a language variety that is not used in the education system.
	9 <i>Lahjeh</i> (accent) is a language variety that is not used in the education system.
	10 <i>Lahjeh</i> (accent) is a language variety which, according to the government, has no official status.
	11 There should be a linguistic label between <i>zabān</i> (language) and <i>guyesh</i> (dialect) to account for all the language varieties that are neither of these two language varieties.
C4-MI. Mutual Intelligibility	12 If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate <i>zabān</i> (languages).
	13 I cannot understand a language variety if I am not familiar with its words.
	14 If someone speaks to me in a <i>lahjeh</i> (accent) of my mother tongue which is different from my own <i>lahjeh</i> (accent), it is always possible to understand them quite well.

	15	If someone speaks to me in a <i>guyesh</i> (dialect) of my mother tongue which is different from my own <i>guyesh</i> (dialect), it is always possible to understand them quite well.
C5-SW. Shared Words	16	If two language varieties share less than 50% of their words, they are generally separate <i>zabān</i> (languages).
	17	If two language varieties share more than 80% of their words, they are generally <i>lahjeh</i> (accents) of the same <i>zabān</i> (language).
	18	If two language varieties share more than 80% of their words, they are generally <i>guyesh</i> (dialects) of the same <i>zabān</i> (language).
C6-WS. Writing System	19	If a language variety has a standardized writing system, it may not be a <i>zabān</i> (language).
	20	<i>Guyesh</i> (dialects) do not have standardized writing systems.
	21	To be considered a <i>zabān</i> (language), a language variety needs to have a standardized writing system.

Moreover, to ensure the representativeness of the questionnaire and to gather more data with heterogeneity, it was piloted two more times after the first round with a population from various speech communities to investigate variation in the way folk express their opinions about different types of language varieties. These pilots helped with analysis of components in light of newly collected data to accommodate folk perspectives in the wording and structure of the questionnaire.

There were multiple underlying reasons for conducting both the second and third pilot iterations. First, the refined version of the questionnaire (see Table 21) was filled out by participants more than half of whom (56.7%) had a linguistics background. Second, due to limited access to Iranian participants at the time, the questionnaire was administered among a limited number of speech communities. As such, the achieved results were not representative of all Iranians' perspectives on language. The study might have yielded different results had it benefited from having various participants with diverse backgrounds, both non-literate and literate. Moreover, considering that speech communities have different priorities in addressing the issues

of language identification, I examined whether using the questionnaire yielded consistent results from the first pilot iteration to the next one.

One of the major reasons for running further pilots was regarding whether the items agreed upon in the final panel session following the first pilot, addressed the issues of language identification. In pilot studies where researchers have access to sufficient number of participants, high factors loadings (correlations) of the items included under each construct depict whether the measurement tool captures what it intended to measure (e.g., see Hinkin, 1998). If an item correlates poorly to a component, it underlines that component as insignificant in addressing the issue under study.

In this study, factorability of data was not analyzed until after the pilot iterations and in the validation process of the questionnaire. This is to say that due to the very low number of participants in the first pilot iteration, I could not investigate the factor loadings (correlations) of all items to their respective construct (i.e., components of the language identification questionnaire). Had I had the opportunity to conduct the first pilot with more participants, I would have been able to statistically evaluate, based on the participants' perspectives, whether the items were representative of how folk define and differentiate different types of language varieties. As such, the second and third pilot iterations and the following expert panel discussions helped identify items which were, to the best of our knowledge, necessary to include. In the case of items on which the expert panel could not reach an agreement, I decided to keep them until later when the factor loadings were established.

Unlike the first pilot which was completed with 30 participants, the second and third pilots were completed with 60 and 70 participants, respectively. Despite minor discrepancies between the first and second pilot results, I decided to keep all the items, even those which generated inconsistent results, until after the third pilot was completed. Analysis of data obtained from the second and third pilot iterations established status of language varieties, the extent of mutual intelligibility and shared words among varieties as well as a standardized writing system among influential parameters in defining and differentiating language varieties with respect to folk perspectives.

In line with the initial pilot results, participants from both the second (see Table 29) and third (see Table 30) pilots expressed mixed views regarding language identification in terms of geography (C1-GE) and number of speakers (C2-NS). A noticeable result among the three pilot iterations was regarding C1.3, that people in cities generally speak *zabān*. The majority (77%) of the first pilot participants disagreed with this statement. In the following pilot iterations, an average of 51% of the participants disagreed with it. Such high extents of disagreement could be due to the features of modern societies. This is to say that “migration, government devolution and technology” (Evans, 2016, p. 55) bring together many culturally and socially diverse groups and renders the idea of a dichotomous urban vs. rural residency obsolete (Lichter & Brown, 2011). This, in turn, may influence the way people identify and use language varieties in urban and rural areas. As interview results indicated, the speech of those who reside in urban areas mirrors features of different types of language varieties. Further, Lichter and Brown (2011) argue that regardless

of the underlying reasons for the flow of people between rural and urban areas, this movement is a mutual one where the flow is not directed from the urban towards the rural areas.

Table 29. Second pilot iteration: C1-GE (Geography) and C2-NS (Number of Speakers) Likert-scale responses

	Agree	Disagree	Don't know
C1.1. People in villages generally speak using <i>guyesh</i> (dialects).	43.3%	43.3%	13.3%
C1.2. People in villages generally speak with <i>lahjeh</i> (accents).	45.0%	40.0%	15.0%
C1.3. People in cities generally speak <i>zabān</i> (languages).	48.3%	51.7%	0.0%
C2.1. A <i>guyesh</i> (dialect) is always part of a larger <i>zabān</i> (language) because it has fewer speakers.	43.3%	46.7%	10.0%
C2.2. If a language variety has only a few thousand speakers, it should generally be considered a <i>guyesh</i> (dialect).	30.0%	50.0%	20.0%
C2.3. If a language variety has at least a million speakers, it should generally be considered a <i>zabān</i> (language).	25.0%	51.7%	23.3%

Table 30. Third pilot iteration: C1-GE (Geography) and C2-NS (Number of Speakers) Likert-scale responses

	Agree	Disagree	Don't know
C1.1. People in villages generally speak using <i>guyesh</i> (dialects).	47.1%	37.1%	15.7%
C1.2. People in villages generally speak with <i>lahjeh</i> (accents).	32.9%	50.0%	17.1%
C1.3. People in cities generally speak <i>zabān</i> (languages).	44.3%	51.4%	4.3%
C2.1. A <i>guyesh</i> (dialect) is always part of a larger <i>zabān</i> (language) because it has fewer speakers.	41.4%	47.1%	11.4%
C2.2. If a language variety has only a few thousand speakers, it should generally be considered a <i>guyesh</i> (dialect).	42.9%	38.6%	18.6%
C2.3. If a language variety has at least a million speakers, it should generally be considered a <i>zabān</i> (language).	28.6%	51.4%	20.0%

As shown in pilot iterations two (see Table 31) and three (see Table 32), analysis of responses across the three iterations further underlined C3.2 as a statement where the results were not entirely compatible. An average of 64% of the participants agreed that *guyesh* is a language variety which is not used in the education system. Pilot one results, however, showed that a population of 37% equally agreed and disagreed with this statement. Such discrepancies in the

participants' perspectives regarding the use of *guyesh* in the education system could be related to a long-standing debate as to whether the language varieties in which students speak should be the medium of knowledge. Regarding this, article 15 of the Constitution of the Islamic Republic of Iran reads that “the use of regional and tribal languages in the press and mass media, as well as for teaching of their literature in schools, is allowed in addition to Persian (Farsi)” (Sheyholislami, 2019, p. 108). Nonetheless, motivated by the ‘one country, one nation, one language’ motto, in Iran, the official language of the country, Farsi, is the medium of education, curricula and textbooks (Asgharzadeh, 2007; Boroujerdi, 1998; Mordai, 2019) even in contexts in which teachers and students are from the same minority language variety (Rust, 2012).

A second statement where responses were not consistent from one pilot to the next was regarding C3.4, that *lahjeh* is a language variety which, according to the government, has no official status. While the population of the participants who agreed was approximately consistent from one iteration to the next, there was a gap between the population of those who disagreed and chose ‘don’t know’ across the pilot iterations (see Tables 31 and 32). Lack of an official status for different types of language varieties other than Farsi underlines the implemented regulations which focus extensively on the expansion, promotion and dissemination of Farsi as the only official language of the country (Moradi, 2019; Sheyholislami, 2012). Such central status given to Farsi may relieve other language varieties of ‘status’ (see Hassanpour, 1992). Despite lack of consistency between the results of the first and following pilots, after the panel discussion session, the respective items (C3.2 and C3.4) were included to further investigate during the Exploratory Factor Analysis (EFA) phase.

Table 31. Second pilot iteration: C3-ST (Status) Likert-scale responses

	Agree	Disagree	Don't know
C3.1. <i>Guyesh</i> (dialect) is a language variety which, according to the government, has no official status.	75.0%	10.0%	15.0%
C3.2. <i>Guyesh</i> (dialect) is a language variety that is not used in the education system.	66.7%	25.0%	8.3%
C3.3. <i>Lahjeh</i> (accent) is a language variety that is not used in the education system.	25.0%	66.7%	8.3%
C3.4. <i>Lahjeh</i> (accent) is a language variety which, according to the government, has no official status.	56.7%	11.7%	31.7%
C3.5. There should be a linguistic label between <i>zabān</i> (language) and <i>guyesh</i> (dialect) to account for all the language varieties that are neither of these two language varieties.	38.3%	43.3%	18.3%

Table 32. Third pilot iteration: C3-ST (Status) Likert-scale responses

	Agree	Disagree	Don't know
C3.1. <i>Guyesh</i> (dialect) is a language variety which, according to the government, has no official status.	68.6%	15.7%	15.7%
C3.2. <i>Guyesh</i> (dialect) is a language variety that is not used in the education system.	57.1%	35.7%	7.1%
C3.3. <i>Lahjeh</i> (accent) is a language variety that is not used in the education system.	30.0%	62.9%	7.1%
C3.4. <i>Lahjeh</i> (accent) is a language variety which, according to the government, has no official status.	57.1%	35.7%	7.1%
C3.5. There should be a linguistic label between <i>zabān</i> (language) and <i>guyesh</i> (dialect) to account for all the language varieties that are neither of these two language varieties.	30.0%	38.6%	31.4%

In component four of the questionnaire (C4-MI), results were consistent for the most part among the three pilots. Statement C4.4 generated inconsistent results with respect to the population who disagreed and chose ‘don’t know’. In the first pilot, 40% chose ‘don’t know’ in C4.4, that if someone speaks to them in a *guyesh* of their mother tongue which is different from their own *guyesh*, it is always possible to understand them quite well. In the following pilots, however, this population was an average of 12% (see Tables 33 and 34). Furthermore, almost 7% of the first pilot participants agreed with C4.4. This population increased to nearly 31% in the second and third pilot iterations. As such, unlike *lahjeh* which was underlined as an intelligible variety type by the interviewees and pilot iteration participants, data did not yield consistent results in identifying *guyesh* in terms of mutual intelligibility. As mentioned before, mutual intelligibility as a stand-alone parameter cannot identify the types of language varieties, since this parameter

correlates significantly with other non-linguistic criteria that are challenging to investigate directly, such as level of exposure to other language varieties (Gooskens, 2007) and attitudes towards language varieties (Mirshahidi, 2017; Mohammadbakhsh et al., 2012). Thus, future qualitative studies can further investigate how one defines and distinguishes *guyesh* from other variety types.

Table 33. Second pilot iteration: C4-MI (Mutual Intelligibility) Likert-scale responses

	Agree	Disagree	Don't know
C4.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate <i>zabān</i> (languages).	68.3%	15.0%	16.7%
C4.2. I cannot understand a language variety if I am not familiar with its words.	73.3%	21.7%	5.0%
C4.3. If someone speaks to me in a <i>lahjeh</i> (accent) of my mother tongue which is different from my own <i>lahjeh</i> (accent), it is always possible to understand them quite well.	80.0%	11.7%	8.3%
C4.4. If someone speaks to me in a <i>guyesh</i> (dialect) of my mother tongue which is different from my own <i>guyesh</i> (dialect), it is always possible to understand them quite well.	31.7%	55.0%	13.3%

Table 34. Third pilot iteration: C4-MI (Mutual Intelligibility) Likert-scale responses

	Agree	Disagree	Don't know
C4.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate <i>zabān</i> (languages).	68.6%	17.1%	14.3%
C4.2. I cannot understand a language variety if I am not familiar with its words.	77.1%	18.6%	4.3%
C4.3. If someone speaks to me in a <i>lahjeh</i> (accent) of my mother tongue which is different from my own <i>lahjeh</i> (accent), it is always possible to understand them quite well.	81.4%	11.4%	7.1%
C4.4. If someone speaks to me in a <i>guyesh</i> (dialect) of my mother tongue which is different from my own <i>guyesh</i> (dialect), it is always possible to understand them quite well.	31.4%	57.1%	11.4%

In component five (C5-SW), 73% of the participants in the first pilot disagreed with C5.3, that two language varieties which share more than 80% of their words are generally *guyesh* of the same *zabān*. This number reduced to an average of 53% in pilots two and three. It is noteworthy that none of the participants from the first pilot agreed with C5.3 while an average of 32.5% expressed agreement in the following pilot iterations.

Table 35. Second pilot iteration: C5-SW (Shared Words) Likert-scale responses

	Agree	Disagree	Don't know
C5.1. If two language varieties share less than 50% of their words, they are generally separate <i>zabān</i> (languages).	51.7%	15.0%	33.3%
C5.2. If two language varieties share more than 80% of their words, they are generally <i>lahjeh</i> (accents) of the same <i>zabān</i> (language).	63.3%	15.0%	21.7%
C5.3. If two language varieties share more than 80% of their words, they are generally <i>guyesh</i> (dialects) of the same <i>zabān</i> (language).	35.0%	53.3%	11.7%

Table 36. Third pilot iteration: C5-SW (Shared Words) Likert-scale responses

	Agree	Disagree	Don't know
C5.1. If two language varieties share less than 50% of their words, they are generally separate <i>zabān</i> (languages).	45.7%	25.7%	28.6%
C5.2. If two language varieties share more than 80% of their words, they are generally <i>lahjeh</i> (accents) of the same <i>zabān</i> (language).	62.9%	18.6%	18.6%
C5.3. If two language varieties share more than 80% of their words, they are generally <i>guyesh</i> (dialects) of the same <i>zabān</i> (language).	32.9%	52.9%	14.3%

Regarding component six (C6-WS), across the three pilot iterations, the participants consistently expressed varied perspectives in identifying a language variety with respect to a writing system except for C6.3. That is, the majority of participants agreed that to be considered a *zabān*, a variety needs to have a standardized writing system (see Table 37 for pilot two and Table 38 for pilot three iteration results).

Table 37. Second pilot iteration: C6-WS (Writing system) Likert-scale responses

	Agree	Disagree	Don't know
C6.1. If a language variety has a standardized writing system, it may not be a <i>zabān</i> (language).	46.7%	15.0%	38.3%
C6.2. <i>Guyesh</i> (dialects) do not have standardized writing systems.	53.3%	35.0%	11.7%
C6.3. To be considered a <i>zabān</i> (language), a language variety needs to have a standardized writing system.	86.7%	10.0%	3.3%

Table 38. Third pilot iteration: C6-WS (Writing system) Likert-scale responses

	Agree	Disagree	Don't know
C6.1. If a language variety has a standardized writing system, it may not be a <i>zabān</i> (language).	42.9%	20.0%	37.1%
C6.2. <i>Guyesh</i> (dialects) do not have standardized writing systems.	52.9%	31.4%	15.7%
C6.3. To be considered a <i>zabān</i> (language), a language variety needs to have a standardized writing system.	85.7%	11.4%	2.9%

Although responses were not entirely consistent from the first to the following iterations, in all of the components the questionnaire generated consistent results from the second pilot iteration to the third one.

4.8. Phase two- Step ten: Validation

I initiated the validation process based on Hinkin (1998) and Neuman and Robson (2012). That said, face and content validity had already been established via multiple sessions with the expert panel. This process led to a questionnaire with six components and 21 items. In order to establish construct validity of the questionnaire, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were run in two separate administration occasions with 203 and 389 participants, respectively (for participants see Section 2.3.1.3.). Nevertheless, there was no point in doing factor analysis without establishing the factorability of data first. Thus, to determine the suitability of data for factor analysis, significant criteria including an appropriate sample size and inter-correlations among items were met (Hinkin, 1998).

With respect to sample size, it is recommended that researchers adopt an item-response ratio of 1 to 4 to 1 to 10 (Hinkin, 1998; Neuman & Robson, 2012). Hinkin (1998) also encourages a conservative approach of 150 observations for EFA provided that item inter-correlations are strong. Hinkin further recommends a minimum sample size of 200 participants for CFA. For these reasons, in this phase, 203 participants for EFA and 389 participants for CFA purposes completed the questionnaire.

Regarding inter-correlations among items, I examined the correlation matrix of the measured variables to ensure correlations in excess of .40 (Field, 2018; Kim & Mueller, 1974;

Tabachnick & Fidell, 2013) and further referred to results from Bartlett’s test, as well as the Kaiser-Meier-Olkin Measure of Sampling Adequacy (i.e., KMO MSA) (see Dziuban & Shirkey, 1974; Hinkin, 1998). A significant Bartlett’s test is considered an indication that it is appropriate to factor analyze the matrix. A KMO MSA $<.50$ indicates that a matrix is unacceptable or as Kaiser and Rice (1974) assert, ‘miserable’ for factoring. As shown in Table 39, a KMO measure of .78 and a significant Bartlett’s test of sphericity ($p=.000$) indicated that there were significant factors to be extracted.

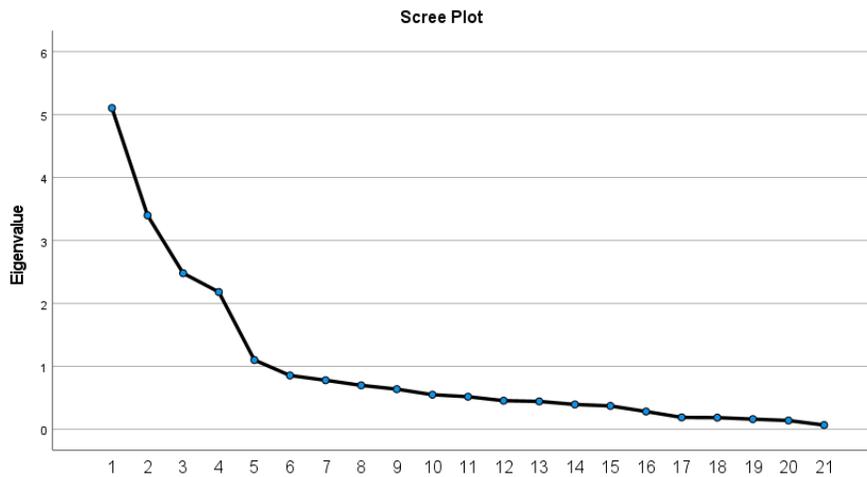
Table 39. KMO and Bartlett’s test results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.78
Bartlett's Test of Sphericity	Approx. Chi-Square	2413.92
	df	210
	Sig.	.000

Having established the factorability of data, based on Principal Component Analysis (PCA), I ran factor analysis to determine the number of factors that accounted for the correlations among the variables. In this regard, both the eigenvalue cut off rule and scree test are identified as appropriate measures (e.g., see Hinkin, 1998; Pituch & Stevens, 2016). As Hinkin (1998) asserts, “if the items have been carefully developed, the number of factors that emerge on both Kaiser and scree criteria should equal the number of scales being developed” (p. 112). For this questionnaire, the scree plot (Figure 1) indicated 5 factors above eigenvalue 1 (Kaiser’s criterion). Following Hinkin’s recommendation in developing reasonably independent scales, varimax orthogonal rotation was chosen for analysis purposes. The generated factors accounted for 67.94% of the total

variance which is above the minimum acceptable target level of 60% (see Hinkin, 1998). These factors accounted for 24.30%, 16.19%, 11.81%, 10.40%, and 5.24% of variance, respectively.

Figure 1. Scree plot with 5 components above eigenvalue 1



I further examined the communality statistics to determine the proportion of variance explained by each item and retain those with higher communalities. All communalities' values were greater than .40. These values ranged between .44 and .86. Based on Principal Component Analysis (PCA), factor analysis results (see Table 40) showed that variables 1 to 6 loaded unambiguously onto component one. Although these variables reflected two parameters, 'geography' and 'number of speakers', they grouped together based on Principal Component Analysis (PCA) results. Variables 7 to 11 reflected 'status' and loaded onto component two. Variables 12 to 15 reflected 'mutual intelligibility' and loaded onto component three. The variables of components four and five reflected 'shared words' and 'writing system', respectively. The one cross-loading variable, V21, did not cause any issue, since the respective variable loaded higher

on component five than four. Since all factor loadings were above .40, at this stage of questionnaire validation, all of the items were kept (see Hinkin, 1998; Pituch & Stevens, 2016).

Table 40. Factor loadings based on Principal Component Analysis (PCA)

	Component				
	1	2	3	4	5
V1	.884				
V2	.880				
V3	.838				
V4	.831				
V5	.764				
V6	.644				
V7		.767			
V8		.763			
V9		.759			
V10		.739			
V11		.712			
V12			.903		
V13			.807		
V14			.769		
V15			.660		
V16				.911	
V17				.891	
V18				.680	
V19					.836
V20					.829
V21				.453	.754

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Before conducting Confirmatory Factor Analysis (CFA) for validity purposes, internal consistency reliability of minimum .70 is an absolute necessity for newly developed measures, though not a sufficient condition (Hinkin, 1998; Nunnally, 1976). As such, coefficient alpha of .82 (see Table 41) indicated strong item covariance.

Table 41. Reliability index based on the questionnaires collected during the Exploratory Factor Analysis (EFA)

Cronbach's Alpha	N of Items
.822	21

Completing the above-mentioned steps was an indication that the questionnaire was internally consistent and had content validity (Hinkin, 1998). To determine whether the items demonstrated goodness of fit, assessed the factor structure and generally confirmed if prior analyses were conducted appropriately, I ran Structural Equation Modelling (SEM) as a confirmatory technique. SEM evaluated the relationship between the latent constructs (5 components of the questionnaire) and measured variables (items of the questionnaire).

To validate the questionnaire, I distributed it among 389 participants (see Section 2.3.1.3.). To investigate the fitness of this questionnaire based on data collected from the respective participants, I used Analysis of a Moment Structures (AMOS) software, version 27. AMOS is a module of SPSS and is used to run Structural Equation Modelling (SEM), Confirmatory Factor Analysis (CFA) and path analysis.

Due to challenges in validating a questionnaire with demographic variables, this section was kept separate. Although chi-square allows for assessing the fitness, researchers have been advised to use it cautiously (e.g., see Byrne, 2010). Increased attention has been focused on goodness of fit indices such as Comparative Fix Index (CFI), Relative Non-centrality Index (RNI), Tucker-Lewis Index (TLI), and Incremental fit index (IFI). These indices may be the most

appropriate to determine the quality of a questionnaire (Byrne, 2010; Medsker et al., 1994; Schumacker & Lomax, 2016). For these indices, values $\geq .95$ are considered as evidence of superior fit and values $\geq .90$ indicate acceptable fit (Byrne, 2010; Whittaker, 2016).

In this phase, I reported both chi-square goodness of fit results (see Table 42) and absolute fit indices (see Table 33). I used CMIN, the traditional approach to testing a questionnaire for goodness of fit (see Kline, 2016). As Table 42 shows, the minimum discrepancy per degree of freedom (CMIN/DF) is 2.21 which is lower than 3, the recommended threshold to deem a questionnaire fit during validation process. As for the more recommended goodness of fit indices, all values were $\geq .90$ and consistent with conventional thresholds for an acceptable fit (see Table 43).

Table 42. Model fit summary: CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	52	396.16	179	.000	2.21
Saturated model	231	.000	0		
Independence model	21	4267.75	210	.000	20.32

Table 43. Model fit summary: Comparative fit indices

Model	NFI Delta1	IFI Delta2	TLI rho2	CFI
Default model	.90	.94	.93	.94
Saturated model	1.00	1.00		1.00
Independence model	.000	.000	.000	.000

Similar to other fit indices, Goodness of fit index (GFI) value of greater than .90 is indicative of an acceptable fit (Pituch & Stevens, 2016). As such, for this study, GFI=.91 was

acceptable. Furthermore, Structural Equation Modelling (SEM) results showed an acceptable value for Root Mean-Square Error of Approximation (RMSEA) (see Table 44). RMSEA is an ‘absolute fit index’, with 0 indicating the ‘best fit’ whereas values > 0 suggest worse fit (Kline, 2016). Although values of .05 or below are indicative of a close-fitting model, values of up to .08 are still acceptable (Whittaker, 2016). As such, the RMSEA = .056 indicated an acceptable fit.

Table 44. RMSEA

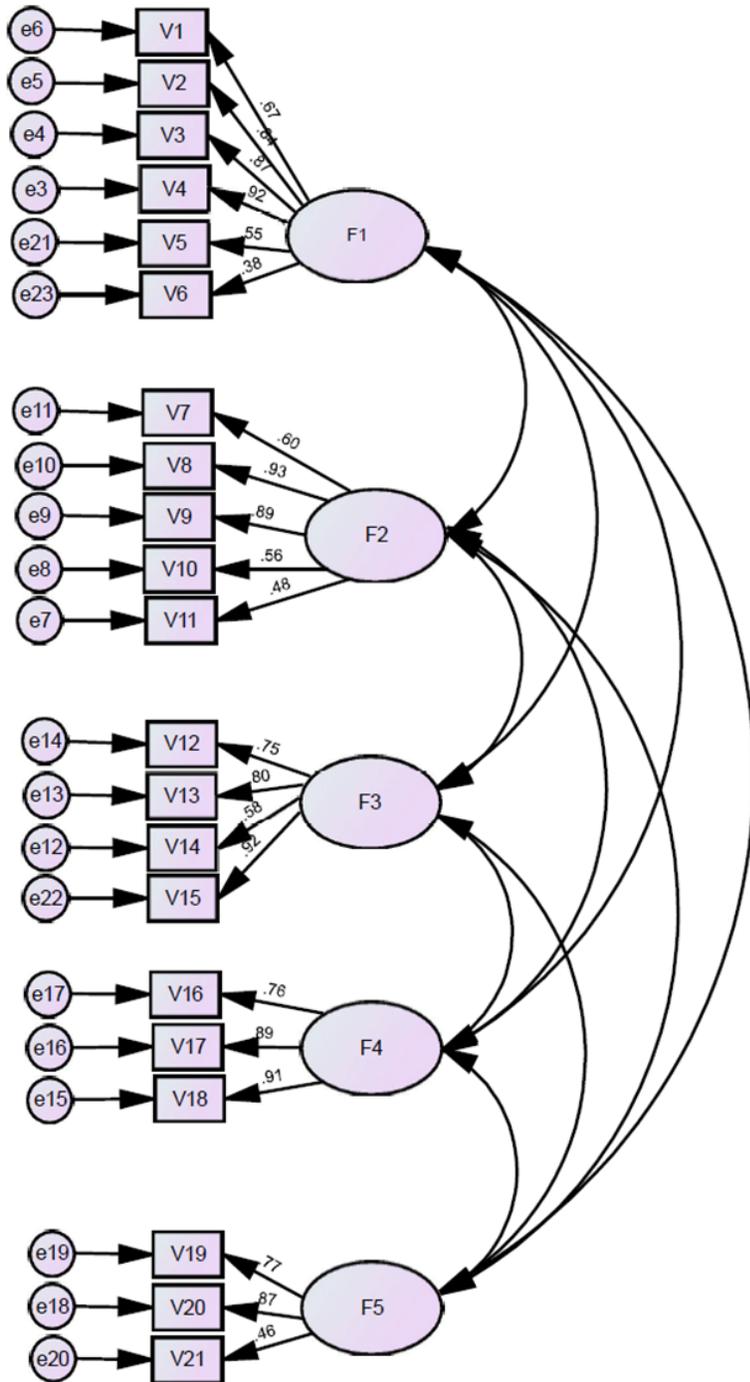
Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.056	.049	.063	.093
Independence model	.223	.217	.229	.000

In order to establish convergent and discriminant validity as a necessary part of determining construct validity (Hinkin, 1998), I analyzed maximum likelihood estimates (see Table 45) including standardized regression weights and correlations. To establish convergent validity, Average Variance Extracted (AVE) values must be greater than .50 for individual components. For the components of this questionnaire, the AVE values were within the acceptable range. Furthermore, the AVE square roots of components one to five were .73, .71, .77, .85, and .72, respectively. These values which were greater than latent construct correlations indicated discriminant validity of the questionnaire. Moreover, according to CR values, all components showed acceptable reliability indices greater than .70. As these values suggest, there are no concerns regarding the validity and reliability of this questionnaire. Finally, Figure 2 illustrates the final language identification questionnaire with path coefficients from each latent construct to its variables.

Table 45. Summary of maximum likelihood estimates to establish convergent and discriminant validities

	CR	AVE	MSV	MaxR(H)	Mutual Intelligibility	Geography and Number of speakers	Status	Shared Words	Writing System
Mutual Intelligibility	0.85	0.60	0.12	0.90	0.77				
Geography and Number of speakers	0.86	0.53	0.02	0.92	0.05	0.73			
Status	0.83	0.51	0.06	0.92	0.07	0.09	0.71		
Shared Words	0.89	0.73	0.24	0.90	0.34	0.15	0.24	0.85	
Writing System	0.75	0.52	0.24	0.82	0.13	0.07	0.08	0.49	0.72

Figure 2. Final language identification questionnaire with 5 components identified during the EFA



5. Hybrid thesis: Part two (Phase three)

Having treated the validation of the questionnaire (see Section 4.8.) in the prior phase with an overall of 592 participants, in **Step eleven**, I administered the questionnaire among a new sample population to investigate the predictive power of folk demographic information in their linguistic perspectives in an attempt to address the research questions. It was administered among a population of 192 participants (see Section 2.3.1.4.). Data analysis revealed significant results regarding folk linguistic perspectives and the associated parameters. Further, regression analysis revealed that in language identification, participants' education, prior formal knowledge of linguistics and mother tongue predicted 53% of the variation in folk linguistic perspectives among the participant sample. The results of this phase are presented in a paper entitled "An analysis of folk linguistic perspectives on language identification in Iran". This paper was submitted to the journal *International Association of Applied Linguistics (AILA)* and is currently under consideration.

5.1. Phase three- Step eleven: Administration of the validated questionnaire

After treating the validation of the questionnaire separately, 192 questionnaires were collected from Farsi, Kurdish, Turkic, and Mazandarani speakers. The first challenge in choosing the proper analytical technique to analyze data gathered via the questionnaire was whether to treat questionnaire data as ordinal or interval. Questionnaire responses are technically considered ordinal data, since data is ranked but the distance between the ranked responses is not equal.

Treating ordinal scales as interval has been among the most controversial quantitative analysis issues (e.g., see Carifio & Perla, 2008; Jamieson, 2004; Norman, 2010). Many researchers consider employing parametric analysis an inappropriate statistical approach to Likert-scale data analysis (Jamieson, 2004). Nonetheless, scholars believe that even when statistical assumptions (e.g., normal distribution of data) are violated, treating Likert-scale data as interval and adopting parametric tests tends to establish equally meaningful results (see Sullivan & Artino, 2013).

Considering the relatively modest sample size (N=192), as well as the questionnaire's employment of Likert-scale rather than Likert type items (see Clason & Dormody, 1994), I analyzed questionnaire data as interval. In addition, parametric data analysis techniques are believed to be more powerful than non-parametric ones (Allen & Seamen, 2007; Boone & Boone, 2012; Sullivan & Artino, 2013).

To continue investigating Iranian folk perspectives about different types of language varieties (research question #1) and examine whether the participants' demographics were among influential parameters in language identification (research question #2), two main measures were taken. After the compilation of questionnaires, as the first measure, I ran and analyzed descriptive

statistics to shed light on the participants' extent of agreement with each of the 21 statements included in the questionnaire. After establishing descriptive statistics, as the second measure I employed hierarchical multiple regression analysis. Regression analysis established the predictive power of participants' demographics (independent variables) in the amount of variation in their overall scores in the questionnaire (dependent variable). Chi-square goodness of fit indices established which demographic categories caused variation in the participants' perspectives about language varieties. This test analysis validated a model with three demographic categories influential in folk perspectives regarding language identification.

In reporting the results here, each component section starts with overall results regarding the participants' extent of agreement with the included statements. Representations of the respective Likert-scale responses via tables assist with data interpretation. The last part focuses on the participants' demographic categories (i.e., mother tongue, formal linguistic knowledge, education, age, and gender) one at a time to further provide details.

5.1.1. C1-GN (Geography and Number of speakers)

5.1.1.1. Overall results

In the validated questionnaire, C1-GN includes six statements which seek participants' perspectives on the relevance of geography (urban vs. rural) and number of speakers in language identification. Whereas 31% of participants agreed, 53% disagreed with these statements. Table 46 further shows a higher level of uncertainty ('don't know' responses) regarding the statements related to number of speakers, as compared to the geography-related statements.

Evans’s (2016) findings, however, do not share the majority participants’ perspectives. In her study, Evans (2016) provided her respondents with a map of Washington state and asked them to draw lines around the areas where they believed people sounded different. She further asked the participants to write down what they would call the way people talked in those areas. The overall results of her study indicated that the dichotomy of urban vs. rural was explicitly used by the majority of respondents. The respondents used labels such as ‘city’ and ‘country’ for different areas, suggesting that the dichotomy was meaningful to them.

Table 46. C1-GN (Geography and Number of speakers): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C1.1. People in villages generally speak using guyesh (dialects).	6.8%	24.0%	19.8%	37.5%	12.0%
C1.2. People in villages generally speak with lahjeh (accents).	7.8%	24.0%	17.7%	35.9%	14.6%
C1.3. People in cities generally speak zabān (languages).	11.5%	28.1%	15.1%	38.0%	7.3%
C1.4. A guyesh (dialect) is always part of a larger zabān (language) because it has fewer speakers.	6.3%	19.3%	11.5%	43.8%	19.3%
C1.5. If a language variety has only a few thousand speakers, it should generally be considered a guyesh (dialect).	4.7%	22.4%	11.5%	43.2%	18.2%
C1.6. If a language variety has at least a million speakers, it should generally be considered a zabān (language).	6.3%	27.1%	15.1%	32.8%	18.8%

As for this research project, the overall results were more in line with Lichter and Brown’s (2011) findings which render the urban/rural dichotomy irrelevant. Nonetheless, as Evans (2016) argues, the majority of folk applications of dialectology in the United States seek respondents’ perspectives using the entire map of the country. Should researchers investigate folk perspectives regarding the map of a single state, results may differ. This perspective brings to attention the significance of conducting qualitative studies which investigate whether folk believe that the distinction between urban and rural areas is meaningfully contributing to language identification. It is also noteworthy that those who disagreed with the geography-related statements were only a fraction of their speech communities. Thus, to decide realistically whether the features of the

speech of urban vs. rural areas are reflected in any of the linguistic terms requires a significant amount of information collected from a greater number of speech communities.

Furthermore, participants' level of disagreement was highest in C1.1, C1.4 and C1.5 all of which identified *guyesh* in terms of the places where it is spoken and number of people who speak it (see Table 46). In line with these perspectives are scholars' arguments (e.g., see Eberhard et al., 2021; Moradi, 2019) which contend that unless resolutions are found to deal with the controversies around what constitutes a language (Anonby et al., 2019) and how to establish the borderlines between language variety types (Anonby & Sabethematabadi, 2019), linguistic endeavours on language identification in terms of this parameter will be ongoing. Among what contributes to this situation are mobile populations and immigration patterns in the country (e.g., see Rezaei et al., 2017). Not only do these features of modern societies make the geographical boundaries blurry, but they may also challenge whether the population of people who speak different types of language varieties in specific areas is comparatively higher or lower than other areas.

5.1.1.2. C1-GN (Geography and Number of speakers): Mother tongue

Almost 55% of Farsi, 47% of Kurdish, 58% of Turkic, and 78% of Mazandarani speakers disagreed with the geography-related statements in this component (see Table 47). An average of 27.5% of Farsi and Kurdish, 34% of Turkic and 16.7% of Mazandarani speakers agreed with the statements related to the number of speakers. Across all language groups, the participants tended to disagree with the relevance of geography and number of speakers as parameters in language identification.

From C1.1 to C1.3, Farsi and Mazandarani speakers of this study disagreed most with C1.2, that people in villages speak with *lahjeh*, whereas Kurdish and Turkic speakers disagreed most with C1.1, that *guyesh* is a variety spoken in villages. Interestingly, a high level of agreement among the four groups was reached about C1.3 in identifying cities as places where people speak *zabān*.

An average of 53% of Farsi, 51% of Kurdish, 48% of Turkic, and 86% of Mazandarani speakers disagreed with C1.4, C1.5 and C1.6 (see Table 47). Among C1.4 to C1.6, for speakers of three language varieties, the highest level of disagreement (Farsi 60.5%, Turkic 52.8%, Mazandarani 91.7%) was with C1.4, that *guyesh* is part of a larger *zabān* due to fewer speakers. Kurdish speakers expressed the highest level of disagreement (60.0%) about C1.5, that a variety with only a few thousand speakers should be considered a *guyesh* (see Table 47).

Table 47. Farsi, Kurdish, Turkic, and Mazandarani speakers' perspectives in C1-GN (Geography and Number of speakers)

		Mother tongue			
		Farsi	Kurdish	Turkic	Mazandarani
C1.1. People in villages generally speak using guyesh (dialects).	Agree	10.5%	5.7%	6.9%	0.0%
	Somewhat agree	28.9%	25.7%	22.2%	8.3%
	Somewhat disagree	15.8%	17.1%	20.8%	41.7%
	Disagree	36.8%	37.1%	38.9%	33.3%
	Don't know	7.9%	14.3%	11.1%	16.7%
C1.2. People in villages generally speak with lahjeh (accents).	Agree	7.9%	4.3%	12.5%	0.0%
	Somewhat agree	15.8%	32.9%	23.6%	0.0%
	Somewhat disagree	15.8%	12.9%	18.1%	50.0%
	Disagree	50.0%	27.1%	36.1%	41.7%
	Don't know	10.5%	22.9%	9.7%	8.3%
C1.3. People in cities generally speak zabān (languages).	Agree	18.4%	7.1%	12.5%	8.3%
	Somewhat agree	31.6%	31.4%	25.0%	16.7%
	Somewhat disagree	13.2%	11.4%	18.1%	25.0%
	Disagree	34.2%	35.7%	41.7%	41.7%
	Don't know	2.6%	14.3%	2.8%	8.3%
C1.4. A guyesh (dialect) is always part of a larger zabān (language) because it has fewer speakers.	Agree	5.3%	5.7%	8.3%	0.0%
	Somewhat agree	15.8%	15.7%	27.8%	0.0%
	Somewhat disagree	15.8%	8.6%	9.7%	25.0%
	Disagree	44.7%	40.0%	43.1%	66.7%
	Don't know	18.4%	30.0%	11.1%	8.3%
C1.5. If a language variety has only a few thousand speakers, it should generally be considered a guyesh (dialect).	Agree	7.9%	2.9%	5.6%	0.0%
	Somewhat agree	23.7%	21.4%	26.4%	0.0%
	Somewhat disagree	10.5%	11.4%	9.7%	25.0%
	Disagree	39.5%	48.6%	37.5%	58.3%
	Don't know	18.4%	15.7%	20.8%	16.7%
C1.6. If a language variety has at least a million speakers, it should generally be considered a zabān (language).	Agree	5.3%	4.3%	8.3%	8.3%
	Somewhat agree	26.3%	31.4%	26.4%	8.3%
	Somewhat disagree	21.1%	10.0%	12.5%	41.7%
	Disagree	26.3%	35.7%	31.9%	41.7%
	Don't know	21.1%	18.6%	20.8%	0.0%

5.1.1.3. C1-GN (Geography and Number of speakers): Formal linguistic knowledge

An average of 55.5% of the participants (see Table 48) who had taken courses in linguistics disagreed with the statements related to geography and number of speakers in C1-GN. Similarly, among those without formal linguistic knowledge, an average of 54.5% and 52% disagreed with the statements related to these two parameters, respectively. The participants with formal linguistic knowledge expressed the highest level of disagreement (58%) with C1.2. With respect to number

of speakers, these participants opposed most (63.2%) to C1.4. On the other hand, those without formal linguistic knowledge disagreed most with C1.1. and C1.5 (58% and 54%). Lastly, an average of 34% of participants, both with and without formal linguistic knowledge, agreed with the geography-related statements (see Table 48).

Table 48. Participants' perspectives with respect to formal linguistic knowledge in C1-GN (Geography and Number of speakers)

		Formal linguistic knowledge	
		No	Yes
C1.1. People in villages generally speak using guyesh (dialects).	Agree	5.8%	10.5%
	Somewhat agree	24.7%	21.1%
	Somewhat disagree	21.4%	13.2%
	Disagree	36.4%	42.1%
	Don't know	11.7%	13.2%
C1.2. People in villages generally speak with lahjeh (accents).	Agree	7.8%	7.9%
	Somewhat agree	24.7%	21.1%
	Somewhat disagree	19.5%	10.5%
	Disagree	33.1%	47.4%
	Don't know	14.9%	13.2%
C1.3. People in cities generally speak zabān (languages).	Agree	9.7%	18.4%
	Somewhat agree	28.6%	26.3%
	Somewhat disagree	16.2%	10.5%
	Disagree	37.0%	42.1%
	Don't know	8.4%	2.6%
C1.4. A guyesh (dialect) is always part of a larger zabān (language) because it has fewer speakers.	Agree	5.2%	10.5%
	Somewhat agree	22.7%	5.3%
	Somewhat disagree	11.0%	13.2%
	Disagree	42.2%	50.0%
C1.5. If a language variety has only a few thousand speakers, it should generally be considered a guyesh (dialect).	Agree	18.8%	21.1%
	Somewhat agree	3.2%	10.5%
	Somewhat disagree	24.7%	13.2%
	Disagree	11.7%	10.5%
	Don't know	42.2%	47.4%
C1.6. If a language variety has at least a million speakers, it should generally be considered a zabān (language).	Agree	18.2%	18.4%
	Somewhat agree	5.2%	10.5%
	Somewhat disagree	30.5%	13.2%
	Disagree	14.3%	18.4%
	Don't know	33.8%	28.9%
	Disagree	33.8%	28.9%
	Don't know	16.2%	28.9%

5.1.1.4. C1-GN (Geography and Number of speakers): Education

MA/PhD holders disagreed slightly more with C1-GN statements compared with BA and college degree holders (see Table 49). An average of 61.5% of MA/PhD, 56.5% of BA and 47.5% of college degree holders disagreed with the statements which underlined urban vs. rural geography. MA/PhD and college participants disagreed most (66.7%) with C1.1 and BA participants disagreed most (56%) with C1.2. Furthermore, an average of 42% of college, 34.7% of BA and 23% of MA/PhD participants agreed with the geography-related statements. Additionally, 52% of college, 50% of BA and 57% of MA/PhD holders disagreed with C1.4, C1.5 and C1.6 while nearly 25% of college and MA/PhD as well as 33% of BA participants agreed with them (see Table 49).

Table 49. College, BA and MA/PhD participants' perspectives regarding C1-GN (Geography and Number of speakers)

		Education		
		College	BA	MA/PhD
C1.1. People in villages generally speak using guyesh (dialects).	Agree	6.1%	9.3%	3.9%
	Somewhat agree	28.8%	25.3%	15.7%
	Somewhat disagree	15.2%	20.0%	25.5%
	Disagree	36.4%	36.0%	41.2%
	Don't know	13.6%	9.3%	13.7%
C1.2. People in villages generally speak with lahjah (accents).	Agree	10.6%	6.7%	5.9%
	Somewhat agree	33.3%	20.0%	17.6%
	Somewhat disagree	12.1%	20.0%	21.6%
	Disagree	31.8%	41.3%	33.3%
	Don't know	12.1%	12.0%	21.6%
C1.3. People in cities generally speak zabān (languages).	Agree	10.6%	14.7%	7.8%
	Somewhat agree	36.4%	28.0%	17.6%
	Somewhat disagree	12.1%	13.3%	21.6%
	Disagree	34.8%	38.7%	41.2%
	Don't know	6.1%	5.3%	11.8%
C1.4. A guyesh (dialect) is always part of a larger zabān (language) because it has fewer speakers.	Agree	6.1%	8.0%	3.9%
	Somewhat agree	24.2%	16.0%	17.6%
	Somewhat disagree	13.6%	10.7%	9.8%
	Disagree	40.9%	45.3%	45.1%
	Don't know	15.2%	20.0%	23.5%
C1.5. If a language variety has only a few thousand speakers, it should generally be considered a guyesh (dialect).	Agree	1.5%	8.0%	3.9%
	Somewhat agree	21.2%	26.7%	17.6%
	Somewhat disagree	12.1%	9.3%	13.7%
	Disagree	42.4%	40.0%	49.0%
	Don't know	22.7%	16.0%	15.7%
C1.6. If a language variety has at least a million speakers, it should generally be considered a zabān (language).	Agree	3.0%	8.0%	7.8%
	Somewhat agree	24.2%	33.3%	21.6%
	Somewhat disagree	12.1%	14.7%	19.6%
	Disagree	34.8%	30.7%	33.3%
	Don't know	25.8%	13.3%	17.6%

5.1.1.5. C1-GN (Geography and Number of speakers): Age

Across all age categories, '40-50' year-old participants disagreed the least with C1.1-C1.6. In this age category, there was not a significant gap in number of those who agreed and disagreed with C1.1-C1.3 (nearly 45% and 48%, respectively). Approximately 57% of other age categories disagreed whereas nearly 33% of these participants agreed with C1.1-C1.3. Nearly, 54% of under

40, 46% of ‘40-50’ and 61% of ‘over 50’ year-old participants opposed to C1.4-C1.6 with number of speakers as the underpinning parameter. Except for ‘40-50’ year-old participants, the rest agreed most with C1.6 (see Table 50).

Table 50. Participants’ perspectives with respect to age in C1-GN (Geography and Number of speakers)

		Age				
		Under 20	20-30	30-40	40-50	Over 50
C1.1. People in villages generally speak using guyesh (dialects).	Agree	7.7%	8.6%	4.9%	3.4%	16.7%
	Somewhat agree	23.1%	18.6%	27.9%	34.5%	0.0%
	Somewhat disagree	19.2%	14.3%	23.0%	24.1%	33.3%
	Disagree	38.5%	45.7%	34.4%	24.1%	33.3%
C1.2. People in villages generally speak with lahjeh (accents).	Don't know	11.5%	12.9%	9.8%	13.8%	16.7%
	Agree	7.7%	7.1%	4.9%	10.3%	33.3%
	Somewhat agree	15.4%	18.6%	31.1%	34.5%	0.0%
	Somewhat disagree	19.2%	12.9%	19.7%	20.7%	33.3%
C1.3. People in cities generally speak zabān (languages).	Disagree	46.2%	42.9%	29.5%	27.6%	16.7%
	Don't know	11.5%	18.6%	14.8%	6.9%	16.7%
	Agree	7.7%	14.3%	9.8%	6.9%	33.3%
	Somewhat agree	34.6%	22.9%	24.6%	44.8%	16.7%
C1.4. A guyesh (dialect) is always part of a larger zabān (language) because it has fewer speakers.	Somewhat disagree	15.4%	7.1%	19.7%	20.7%	33.3%
	Disagree	42.3%	41.4%	39.3%	27.6%	16.7%
	Don't know	0.0%	14.3%	6.6%	0.0%	0.0%
	Agree	11.5%	7.1%	3.3%	6.9%	0.0%
C1.5. If a language variety has only a few thousand speakers, it should generally be considered a guyesh (dialect).	Somewhat agree	11.5%	14.3%	21.3%	34.5%	16.7%
	Somewhat disagree	3.8%	11.4%	11.5%	13.8%	33.3%
	Disagree	57.7%	44.3%	41.0%	37.9%	33.3%
	Don't know	15.4%	22.9%	23.0%	6.9%	16.7%
C1.6. If a language variety has at least a million speakers, it should generally be considered a zabān (language).	Agree	7.7%	5.7%	1.6%	3.4%	16.7%
	Somewhat agree	30.8%	15.7%	21.3%	37.9%	0.0%
	Somewhat disagree	7.7%	11.4%	6.6%	20.7%	33.3%
	Disagree	50.0%	44.3%	50.8%	20.7%	33.3%
C1.6. If a language variety has at least a million speakers, it should generally be considered a zabān (language).	Don't know	3.8%	22.9%	19.7%	17.2%	16.7%
	Agree	7.7%	5.7%	6.6%	3.4%	16.7%
	Somewhat agree	38.5%	24.3%	23.0%	34.5%	16.7%
	Somewhat disagree	11.5%	14.3%	11.5%	24.1%	33.3%
C1.6. If a language variety has at least a million speakers, it should generally be considered a zabān (language).	Disagree	38.5%	30.0%	41.0%	20.7%	16.7%
	Don't know	3.8%	25.7%	18.0%	17.2%	16.7%

5.1.1.6. C1-GN (Geography and Number of speakers): Gender

More female participants opposed the relevance of geography and number of speakers than did male participants in language identification. Nearly 50% of the males and 57% of the females disagreed whereas nearly 33% of both male and female participants agreed with geography-related statements. Female participants opposed most (60.2%) to C1.1 and male participants disagreed most (56.5%) with C1.2 (see Table 51).

Table 51. Male and female perspectives in C1-GN (Geography and Number of speakers)

		Gender	
		Male	Female
C1.1. People in villages generally speak using guyesh (dialects).	Agree	5.8%	7.3%
	Somewhat agree	26.1%	22.8%
	Somewhat disagree	21.7%	18.7%
	Disagree	30.4%	41.5%
	Don't know	15.9%	9.8%
C1.2. People in villages generally speak with lahjeh (accents).	Agree	5.8%	8.9%
	Somewhat agree	18.8%	26.8%
	Somewhat disagree	14.5%	19.5%
	Disagree	42.0%	32.5%
	Don't know	18.8%	12.2%
C1.3. People in cities generally speak zabān (languages).	Agree	11.6%	11.4%
	Somewhat agree	33.3%	25.2%
	Somewhat disagree	13.0%	16.3%
	Disagree	29.0%	43.1%
	Don't know	13.0%	4.1%
C1.4. A guyesh (dialect) is always part of a larger zabān (language) because it has fewer speakers.	Agree	5.8%	6.5%
	Somewhat agree	20.3%	18.7%
	Somewhat disagree	13.0%	10.6%
	Disagree	46.4%	42.3%
	Don't know	14.5%	22.0%
C1.5. If a language variety has only a few thousand speakers, it should generally be considered a guyesh (dialect).	Agree	2.9%	5.7%
	Somewhat agree	14.5%	26.8%
	Somewhat disagree	8.7%	13.0%
	Disagree	34.8%	48.0%
	Don't know	39.1%	6.5%
C1.6. If a language variety has at least a million speakers, it should generally be considered a zabān (language).	Agree	7.2%	5.7%
	Somewhat agree	23.2%	29.3%
	Somewhat disagree	11.6%	17.1%
	Disagree	20.3%	39.8%
	Don't know	37.7%	8.1%

Statements C1.4-C1.6 with number of speakers as underpinning parameter were disagreed with by approximately 45% and 57% of male and female participants, respectively. Nearly 25% of male and 33% of female participants agreed with C1.4-C1.6. Findings showed that C1.4 and C1.5 were most disagreed upon by both male and female participants. The highest agreement was reached about C1.6 (see Table 51).

5.1.2. C2-ST (Status)

5.1.2.1. Overall results

C2 statements investigate the status of different types of language varieties within the educational system and according to the government. The results indicated that most of the participants identified *guyesh* and *lahjeh* as language varieties which have no official status according to the government. Nearly 73% and 50% of the participants respectively identified *guyesh* and *lahjeh* as varieties which are not used in the education system (see Table 52). As May (2006) suggests, specifically minority groups are prone to downplay the significance of their language variety or how they refer to it, simply because they may grow up in a cultural environment where language issues are highly politicized. Minority languages may end up being called dialects or accents of a language not just by the dominant majority but by the minority themselves. This is most likely due to their not being aware of any other status options since this is the ideology in which they grew up.

Considering the available literature on the status of language varieties other than Farsi in educational policies and environments, social media and government, it is not far-fetched to suggest that the participants' linguistic perspectives may reflect a similar situation where not only

the sociopolitical mandates of Iran, but also the participants (i.e., community members) regard language varieties other than the official language of the country as less acceptable and lacking prestige. While similar to *guyesh*, *lahjeh* was more often identified as having no official status, regarding its use in the education system, perspectives were varied. Although pronunciation has always been key in differentiating *lahjeh* from other types of language varieties, the same sociocultural complexities and biases may apply.

Table 52. C2-ST (Status): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C2.1. <i>Guyesh</i> (dialect) is a language variety which, according to the government, has no official status.	27.1%	47.9%	0.5%	7.8%	16.7%
C2.2. <i>Guyesh</i> (dialect) is a language variety that is not used in the education system.	23.4%	49.5%	2.6%	16.7%	7.8%
C2.3. <i>Lahjeh</i> (accent) is a language variety that is not used in the education system.	12.5%	38.0%	4.7%	38.5%	6.3%
C2.4. <i>Lahjeh</i> (accent) is a language variety which, according to the government, has no official status.	22.4%	47.9%	3.6%	12.0%	14.1%
C2.5. There should be a linguistic label between <i>zabān</i> (language) and <i>guyesh</i> (dialect) to account for all the language varieties that are neither of these two language varieties.	14.6%	28.6%	4.2%	12.5%	40.1%

5.1.2.2. C2-ST (Status): Mother tongue

About 75% of Farsi, Kurdish and Turkic as well as nearly 42% of Mazandarani speakers agreed with C2.1 and C2.4. While only 14% of Farsi, Kurdish and Turkic speakers disagreed with C2.4, 50% of Mazandarani speakers disagreed with it. In terms of use, nearly 75% of Farsi, Kurdish and Turkic as well as 58% of Mazandarani speakers agreed with C2.2. Nonetheless, the extent of this agreement was lower in C2.3. That said, 51% of all speakers agreed with C2.3, that *lahjeh* is not used in the education system, whereas 45% disagreed with it. Lastly, 15% of Farsi, Kurdish and Turkic as well as 50% of Mazandarani speakers disagreed with C2.5. More than 50%

of Farsi, an average of 42% of Kurdish and Turkic and 25% of Mazandarani speakers agreed with C2.5 (see Table 53).

Table 53. Farsi, Kurdish, Turkic, and Mazandarani speakers' perspectives in C2-ST (Status)

		Mother tongue			
		Farsi	Kurdish	Turkic	Mazandarani
C2.1. Guyesh (dialect) is a language variety which, according to the government, has no official status.	Agree	15.8%	32.9%	29.2%	16.7%
	Somewhat agree	57.9%	48.6%	45.8%	25.0%
	Somewhat disagree	2.6%	0.0%	0.0%	0.0%
	Disagree	7.9%	5.7%	6.9%	25.0%
	Don't know	15.8%	12.9%	18.1%	33.3%
C2.2. Guyesh (dialect) is a language variety that is not used in the education system.	Agree	13.2%	25.7%	27.8%	16.7%
	Somewhat agree	57.9%	60.0%	36.1%	41.7%
	Somewhat disagree	2.6%	1.4%	2.8%	8.3%
	Disagree	18.4%	8.6%	20.8%	33.3%
	Don't know	7.9%	4.3%	12.5%	0.0%
C2.3. Lahjeh (accent) is a language variety that is not used in the education system.	Agree	7.9%	10.0%	18.1%	8.3%
	Somewhat agree	44.7%	41.4%	30.6%	41.7%
	Somewhat disagree	5.3%	7.1%	1.4%	8.3%
	Disagree	42.1%	32.9%	41.7%	41.7%
	Don't know	0.0%	8.6%	8.3%	0.0%
C2.4. Lahjeh (accent) is a language variety which, according to the government, has no official status.	Agree	18.4%	18.6%	27.8%	25.0%
	Somewhat agree	52.6%	52.9%	45.8%	16.7%
	Somewhat disagree	2.6%	2.9%	4.2%	8.3%
	Disagree	13.2%	11.4%	6.9%	41.7%
	Don't know	13.2%	14.3%	15.3%	8.3%
C2.5. There should be a linguistic label between zabān (language) and guyesh (dialect) to account for all the language varieties that are neither of these two language varieties.	Agree	15.8%	14.3%	15.3%	8.3%
	Somewhat agree	36.8%	31.4%	23.6%	16.7%
	Somewhat disagree	5.3%	1.4%	5.6%	8.3%
	Disagree	13.2%	10.0%	9.7%	41.7%
	Don't know	28.9%	42.9%	45.8%	25.0%

5.1.2.3. C2-ST (Status): Formal linguistic knowledge

An average of 68% of the participants who had taken linguistics courses and 74% of those who had not, agreed with C2.1 and C2.4 (see Table 54). Results revealed that in both groups, the number of those who disagreed with C2.4 is twice the number of those who disagreed with C2.1. Among those without formal linguistic knowledge, findings showed an equal number (an average of 45%) of participants who agreed and disagreed with C2.3. Nonetheless, such finding did not apply to the participants with formal linguistic knowledge. Nearly 66% of the respective participants agreed with C2.3 while 33% of them disagreed with it. Furthermore, the majority in

both groups agreed with C2.2. Lastly, only an average of 16% in both groups agreed with C2.5 whereas 40% disagreed with it.

Table 54. Participants’ perspectives with respect to formal linguistic knowledge in C2-ST (Status)

		Formal linguistic knowledge	
		No	Yes
C2.1. Guyesh (dialect) is a language variety which, according to the government, has no official status.	Agree	28.6%	21.1%
	Somewhat agree	49.4%	42.1%
	Somewhat disagree	0.0%	2.6%
	Disagree	9.1%	2.6%
	Don't know	13.0%	31.6%
C2.2. Guyesh (dialect) is a language variety that is not used in the education system.	Agree	22.1%	28.9%
	Somewhat agree	48.1%	55.3%
	Somewhat disagree	2.6%	2.6%
	Disagree	18.8%	7.9%
	Don't know	8.4%	5.3%
C2.3. Lahjeh (accent) is a language variety that is not used in the education system.	Agree	12.3%	13.2%
	Somewhat agree	35.1%	50.0%
	Somewhat disagree	4.5%	5.3%
	Disagree	40.9%	28.9%
	Don't know	7.1%	2.6%
C2.4. Lahjeh (accent) is a language variety which, according to the government, has no official status.	Agree	22.1%	23.7%
	Somewhat agree	47.4%	50.0%
	Somewhat disagree	3.9%	2.6%
	Disagree	13.0%	7.9%
	Don't know	13.6%	15.8%
C2.5. There should be a linguistic label between zabān (language) and guyesh (dialect) to account for all the language varieties that are neither of these two language varieties.	Agree	14.9%	13.2%
	Somewhat agree	29.9%	23.7%
	Somewhat disagree	3.9%	5.3%
	Disagree	13.0%	10.5%
	Don't know	38.3%	47.4%

5.1.2.4. C2-ST (Status): Education

There was no significant gap among the linguistic perspectives of college, undergraduate and graduate participants. Approximately 71% of all participants agreed with C2.1 and C2.4. The number of participants who disagreed with C2.1 is approximately half of those who disagreed with C2.4 (see Table 55). An average of 73% of all degree holders agreed with C2.2, 50% of them agreed with C2.3 while nearly 33% disagreed with C2.3. Finally, approximately 44% of college,

BA and MA/PhD agreed with C2.5 while 41% of them chose ‘don’t know’ as their response in this regard.

Table 55. College, BA and MA/PhD participants’ perspectives regarding C2-ST (Status)

		Education		
		College	BA	MA/PhD
C2.1. Guyesh (dialect) is a language variety which, according to the government, has no official status.	Agree	30.3%	28.0%	21.6%
	Somewhat agree	45.5%	48.0%	51.0%
	Somewhat disagree	0.0%	0.0%	2.0%
	Disagree	7.6%	10.7%	3.9%
	Don't know	16.7%	13.3%	21.6%
C2.2. Guyesh (dialect) is a language variety that is not used in the education system.	Agree	21.2%	25.3%	23.5%
	Somewhat agree	53.0%	44.0%	52.9%
	Somewhat disagree	1.5%	2.7%	3.9%
	Disagree	13.6%	22.7%	11.8%
	Don't know	10.6%	5.3%	7.8%
C2.3. Lahjeh (accent) is a language variety that is not used in the education system.	Agree	9.1%	18.7%	7.8%
	Somewhat agree	31.8%	40.0%	43.1%
	Somewhat disagree	7.6%	1.3%	5.9%
	Disagree	40.9%	36.0%	39.2%
	Don't know	10.6%	4.0%	3.9%
C2.4. Lahjeh (accent) is a language variety which, according to the government, has no official status.	Agree	24.2%	22.7%	19.6%
	Somewhat agree	48.5%	46.7%	49.0%
	Somewhat disagree	6.1%	2.7%	2.0%
	Disagree	10.6%	13.3%	11.8%
	Don't know	10.6%	14.7%	17.6%
C2.5. There should be a linguistic label between zabān (language) and guyesh (dialect) to account for all the language varieties that are neither of these two language varieties.	Agree	15.2%	14.7%	13.7%
	Somewhat agree	31.8%	24.0%	31.4%
	Somewhat disagree	4.5%	4.0%	3.9%
	Disagree	10.6%	16.0%	9.8%
	Don't know	37.9%	41.3%	41.2%

5.1.2.5. C2-ST (Status): Age

Approximately 75% of the participants in all age categories agreed with C2.1 and C2.4 (see Table 56). Participants in ‘20-30’ age category both agreed and disagreed the least with these statements, whereas ‘over 50’ year-old participants agreed the most with them. Nearly 71% of the participants under 50 years of age agreed with C2.2. Only 50% of ‘over 50’ year-old participants held the same perspective. For this age category, this estimate increased to 67% in C2.3. Lastly, an average of 49% of the participants in all other age categories agreed with C2.3. An estimate of

25% of the participants in all age categories disagreed with C2.2 and 42% of them disagreed with C2.3. Finally, in every age category the population of those who disagreed with C2.5 is at least a quarter and at most half of those who agreed with it.

Table 56. Participants' perspectives with respect to age in C2-ST (Status)

		Age				
		Under 20	20-30	30-40	40-50	Over 50
C2.1. Guyesh (dialect) is a language variety which, according to the government, has no official status.	Agree	34.6%	22.9%	27.9%	27.6%	33.3%
	Somewhat agree	46.2%	45.7%	50.8%	48.3%	50.0%
	Somewhat disagree	0.0%	0.0%	1.6%	0.0%	0.0%
	Disagree	11.5%	4.3%	4.9%	17.2%	16.7%
	Don't know	7.7%	27.1%	14.8%	6.9%	0.0%
C2.2. Guyesh (dialect) is a language variety that is not used in the education system.	Agree	26.9%	21.4%	24.6%	24.1%	16.7%
	Somewhat agree	38.5%	54.3%	54.1%	41.4%	33.3%
	Somewhat disagree	0.0%	4.3%	1.6%	3.4%	0.0%
	Disagree	30.8%	12.9%	8.2%	27.6%	33.3%
	Don't know	3.8%	7.1%	11.5%	3.4%	16.7%
C2.3. Lahjeh (accent) is a language variety that is not used in the education system.	Agree	19.2%	10.0%	13.1%	13.8%	0.0%
	Somewhat agree	26.9%	44.3%	34.4%	34.5%	66.7%
	Somewhat disagree	0.0%	7.1%	1.6%	10.3%	0.0%
	Disagree	46.2%	34.3%	41.0%	37.9%	33.3%
	Don't know	7.7%	4.3%	9.8%	3.4%	0.0%
C2.4. Lahjeh (accent) is a language variety which, according to the government, has no official status.	Agree	26.9%	20.0%	19.7%	31.0%	16.7%
	Somewhat agree	38.5%	48.6%	52.5%	41.4%	66.7%
	Somewhat disagree	3.8%	2.9%	1.6%	10.3%	0.0%
	Disagree	11.5%	10.0%	14.8%	10.3%	16.7%
	Don't know	19.2%	18.6%	11.5%	6.9%	0.0%
C2.5. There should be a linguistic label between zabān (language) and guyesh (dialect) to account for all the language varieties that are neither of these two language varieties.	Agree	15.4%	11.4%	18.0%	13.8%	16.7%
	Somewhat agree	15.4%	31.4%	27.9%	34.5%	33.3%
	Somewhat disagree	3.8%	4.3%	1.6%	10.3%	0.0%
	Disagree	23.1%	7.1%	13.1%	13.8%	16.7%
	Don't know	42.3%	45.7%	39.3%	27.6%	33.3%

5.1.2.6. C2-ST (Status): Gender

An average of 78% of male as well as 69.5% of female participants agreed with C2.1 and C2.4. Results showed that compared with female participants, fewer male participants disagreed with these statements. Nearly 50% of both male and female participants agreed with C2.3 and 44% of these participants disagreed with it. The male population who agreed with C2.2 (85.5%) was almost 20% more than the female population who held the same perspective; 8.7% of male and

25% of female participants disagreed with C2.2. Finally, 38% of male and 46% of female participants agreed with C2.5 (see Table 57).

Table 57. Male and female perspectives in C2-ST (Status)

		Gender	
		Male	Female
C2.1. Guyesh (dialect) is a language variety which, according to the government, has no official status.	Agree	30.4%	25.2%
	Somewhat agree	46.4%	48.8%
	Somewhat disagree	0.0%	0.8%
	Disagree	4.3%	9.8%
	Don't know	18.8%	15.4%
C2.2. Guyesh (dialect) is a language variety that is not used in the education system.	Agree	26.1%	22.0%
	Somewhat agree	59.4%	43.9%
	Somewhat disagree	0.0%	4.1%
	Disagree	8.7%	21.1%
	Don't know	5.8%	8.9%
C2.3. Lahjeh (accent) is a language variety that is not used in the education system.	Agree	8.7%	14.6%
	Somewhat agree	43.5%	35.0%
	Somewhat disagree	7.2%	3.3%
	Disagree	39.1%	38.2%
	Don't know	1.4%	8.9%
C2.4. Lahjeh (accent) is a language variety which, according to the government, has no official status.	Agree	21.7%	22.8%
	Somewhat agree	58.0%	42.3%
	Somewhat disagree	5.8%	2.4%
	Disagree	7.2%	14.6%
	Don't know	7.2%	17.9%
C2.5. There should be a linguistic label between zabān (language) and guyesh (dialect) to account for all the language varieties that are neither of these two language varieties.	Agree	13.0%	15.4%
	Somewhat agree	24.6%	30.9%
	Somewhat disagree	5.8%	3.3%
	Disagree	8.7%	14.6%
	Don't know	47.8%	35.8%

5.1.3. C3-MI (Mutual Intelligibility)

5.1.3.1. Overall results

C3 statements investigate whether mutual intelligibility among speakers of different language varieties influences their perspectives regarding these varieties (see Table 58). An analysis of Likert-scale responses underlined the correlation between the extent of mutual intelligibility and familiarity with words in language identification. The majority of participants confirmed that for varieties to be intelligible, language users need to be familiar with the words of

these language varieties (C3.2). Familiarity with the words, in turn, leads to an expansion of mutual intelligibility among the respective varieties (e.g., see Gooskens, 2007; Gooskens et al., 2018).

Table 58. C3-MI (Mutual Intelligibility): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C3.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate <i>zabān</i> (languages).	37.5%	33.3%	4.7%	16.7%	7.8%
C3.2. I cannot understand a language variety if I am not familiar with its words.	42.7%	34.9%	3.6%	12.5%	6.3%
C3.3. If someone speaks to me in a <i>lahjeh</i> (accent) of my mother tongue which is different from my own <i>lahjeh</i> (accent), it is always possible to understand them quite well.	34.4%	42.2%	4.7%	8.3%	10.4%
C3.4. If someone speaks to me in a <i>guyesh</i> (dialect) of my mother tongue which is different from my own <i>guyesh</i> (dialect), it is always possible to understand them quite well.	17.2%	32.3%	12.0%	13.5%	25.0%

The majority of participants identified two *lahjeh* of the same *zabān* as varieties which are always mutually intelligible (C3.3) whereas two *zabān* are mutually unintelligible (C3.1). Similarly, databases of the world’s languages including *Ethnologue* (Eberhard et al., 2021) and *Glottolog* (Hammarström et al., 2021) posit that a language variety that is not mutually intelligible with other varieties should be counted as a separate language. These findings are further in line with Iranian scholars’ (e.g., see Dabir-Moghaddam, 2008; Modarresi, 2014) views regarding the extent of mutual intelligibility with respect to *lahjeh* and *zabān*. Considering the participants’ level of agreement (77.6%) regarding familiarity with the words of a variety in understanding it, it can be inferred that people are not familiar, to a great extent, with the words of a separate *zabān*, whereas they are familiar with and understand the words of a *lahjeh* which belongs to their mother tongue.

Although mutual intelligibility has been recognized as one of the most common and prominent parameters in language identification (Eberhard et al., 2021; Modarresi, 2014), identification of *guyesh* seemed to be far more complicated for the participants. Criteria such as ‘mutual intelligibility’ or ‘structural similarity’ cannot set boundaries between dialects and other

types of language varieties (Wolfram, 2017). Thus, future research needs to define *guyesh* with regard to a combination of parameters and further distinguish this variety from *lahjeh* and *zabān*. As Coupland (1988) pointed out, it is not individual parameters but rather interrelated parameters which identify a variety.

5.1.3.2. C3-MI (Mutual Intelligibility): Mother tongue

Majority of the speakers (an average of 89.5% of Farsi, 68% of Kurdish, 73% of Turkic and 66% of Mazandarani speakers) agreed with C3.1 and C3.2. Almost 90% of Farsi speakers agreed with C3.3 and less than 50% of them agreed with C3.4. Results further showed that unlike Farsi speakers, only 50% of Mazandarani speakers agreed with C3.3. Nearly 75% of Mazandarani speakers agreed with C3.4. Almost 75% of Kurdish and Turkic speakers agreed with C3.3. More than half (55%) of Kurdish speakers agreed whereas 20% disagreed with C3.4. Among Turkic speakers, an approximately equal population (38%) agreed and disagreed with C3.4.

Table 59. Farsi, Kurdish, Turkic, and Mazandarani speakers' perspectives in C3-MI (Mutual Intelligibility)

		Mother tongue			
		Farsi	Kurdish	Turkic	Mazandarani
C3.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate zabān (languages).	Agree	55.3%	31.4%	33.3%	41.7%
	Somewhat agree	34.2%	30.0%	37.5%	25.0%
	Somewhat disagree	2.6%	4.3%	5.6%	8.3%
	Disagree	5.3%	22.9%	15.3%	25.0%
	Don't know	2.6%	11.4%	8.3%	0.0%
C3.2. I cannot understand a language variety if I am not familiar with its words.	Agree	57.9%	40.0%	38.9%	33.3%
	Somewhat agree	31.6%	35.7%	36.1%	33.3%
	Somewhat disagree	0.0%	4.3%	5.6%	0.0%
	Disagree	5.3%	11.4%	13.9%	33.3%
	Don't know	5.3%	8.6%	5.6%	0.0%
C3.3. If someone speaks to me in a lahjeh (accent) of my mother tongue which is different from my own lahjeh (accent), it is always possible to understand them quite well.	Agree	42.1%	30.0%	37.5%	16.7%
	Somewhat agree	47.4%	45.7%	37.5%	33.3%
	Somewhat disagree	0.0%	5.7%	2.8%	25.0%
	Disagree	2.6%	7.1%	9.7%	25.0%
	Don't know	7.9%	11.4%	12.5%	0.0%
C3.4. If someone speaks to me in a guyesh (dialect) of my mother tongue which is different from my own guyesh (dialect), it is always possible to understand them quite well.	Agree	18.4%	18.6%	13.9%	25.0%
	Somewhat agree	28.9%	37.1%	26.4%	50.0%
	Somewhat disagree	10.5%	8.6%	15.3%	16.7%
	Disagree	5.3%	11.4%	20.8%	8.3%
	Don't know	36.8%	24.3%	23.6%	0.0%

5.1.3.3. C3-MI (Mutual Intelligibility): Formal linguistic knowledge

Most participants, with or without formal linguistic knowledge, agreed with the statements included in this component. Nonetheless, these participants did not reach the same decision regarding C3.4 (see Table 60 below). That is, nearly 50% of the two groups agreed with C3.4 and 50% either disagreed with it or chose 'don't know'.

Table 60. Participants' perspectives with respect to formal linguistic knowledge in C3-MI (Mutual Intelligibility)

		Formal linguistic knowledge	
		No	Yes
C3.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate zabān (languages).	Agree	33.1%	55.3%
	Somewhat agree	37.0%	18.4%
	Somewhat disagree	5.8%	0.0%
	Disagree	16.9%	15.8%
	Don't know	7.1%	10.5%
C3.2. I cannot understand a language variety if I am not familiar with its words.	Agree	39.0%	57.9%
	Somewhat agree	39.6%	15.8%
	Somewhat disagree	4.5%	0.0%
	Disagree	11.7%	15.8%
	Don't know	5.2%	10.5%
C3.3. If someone speaks to me in a lahjeh (accent) of my mother tongue which is different from my own lahjeh (accent), it is always possible to understand them quite well.	Agree	31.8%	44.7%
	Somewhat agree	45.5%	28.9%
	Somewhat disagree	5.8%	0.0%
	Disagree	6.5%	15.8%
	Don't know	10.4%	10.5%
C3.4. If someone speaks to me in a guyesh (dialect) of my mother tongue which is different from my own guyesh (dialect), it is always possible to understand them quite well.	Agree	15.6%	23.7%
	Somewhat agree	34.4%	23.7%
	Somewhat disagree	13.0%	7.9%
	Disagree	14.3%	10.5%
	Don't know	22.7%	34.2%

5.1.3.4. C3-MI (Mutual Intelligibility): Education

Almost 66% of the participants with college and graduate degrees and 77% of BA degree holders agreed with C3.1. Approximately 77% of the participants from different educational backgrounds agreed with C3.2. Nearly 83% of college and 73% of undergraduate and graduate participants agreed with C3.3. Finally, 49% of college, BA, and MA/PhD holders agreed with C3.4 (see Table 61).

Table 61. College, BA and MA/PhD participants’ perspectives regarding C3-MI (Mutual Intelligibility)

		Education		
		College	BA	MA/PhD
C3.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate zabān (languages).	Agree	40.9%	38.7%	31.4%
	Somewhat agree	25.8%	38.7%	35.3%
	Somewhat disagree	6.1%	5.3%	2.0%
	Disagree	19.7%	10.7%	21.6%
	Don't know	7.6%	6.7%	9.8%
C3.2. I cannot understand a language variety if I am not familiar with its words.	Agree	48.5%	44.0%	33.3%
	Somewhat agree	25.8%	40.0%	39.2%
	Somewhat disagree	6.1%	1.3%	3.9%
	Disagree	12.1%	12.0%	13.7%
	Don't know	7.6%	2.7%	9.8%
C3.3. If someone speaks to me in a lahjeh (accent) of my mother tongue which is different from my own lahjeh (accent), it is always possible to understand them quite well.	Agree	45.5%	29.3%	27.5%
	Somewhat agree	37.9%	44.0%	45.1%
	Somewhat disagree	4.5%	5.3%	3.9%
	Disagree	4.5%	9.3%	11.8%
	Don't know	7.6%	12.0%	11.8%
C3.4. If someone speaks to me in a guyesh (dialect) of my mother tongue which is different from my own guyesh (dialect), it is always possible to understand them quite well.	Agree	25.8%	16.0%	7.8%
	Somewhat agree	28.8%	33.3%	35.3%
	Somewhat disagree	10.6%	14.7%	9.8%
	Disagree	15.2%	13.3%	11.8%
	Don't know	19.7%	22.7%	35.3%

5.1.3.5. C3-MI (Mutual Intelligibility): Age

Results showed that age did not make a significant difference in how the participants regarded mutual intelligibility in addressing language identification issues. This is to say that an average of 83% of ‘under 20’, 70% of ‘20-30’, 76% of ‘30-40’, 67% of ‘40-50’, and the entire population of ‘over 50’ year-old participants agreed with C3.1 and C3.2. Nearly 80% of 20 to 50-year-old and 61% of ‘under 20’ year-old participants agreed with C3.3. Almost 25% of the participants in every age category disagreed with C3.4. Regarding this, the highest level of disagreement was expressed by the participants in the age category of ‘40-50’. Lastly, a little over 50% of ‘30-40’ year-old participants agreed with C3.4 (see Table 62).

Table 62. Participants' perspectives with respect to age in C3-MI (Mutual Intelligibility)

		Age				
		Under 20	20-30	30-40	40-50	Over 50
C3.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate zabān (languages).	Agree	42.3%	32.9%	39.3%	34.5%	66.7%
	Somewhat agree	38.5%	32.9%	34.4%	27.6%	33.3%
	Somewhat disagree	0.0%	4.3%	3.3%	13.8%	0.0%
	Disagree	15.4%	18.6%	16.4%	17.2%	0.0%
	Don't know	3.8%	11.4%	6.6%	6.9%	0.0%
C3.2. I cannot understand a language variety if I am not familiar with its words.	Agree	38.5%	34.3%	45.9%	51.7%	83.3%
	Somewhat agree	46.2%	40.0%	32.8%	20.7%	16.7%
	Somewhat disagree	3.8%	1.4%	3.3%	10.3%	0.0%
	Disagree	11.5%	15.7%	13.1%	6.9%	0.0%
	Don't know	0.0%	8.6%	4.9%	10.3%	0.0%
C3.3. If someone speaks to me in a lahjeh (accent) of my mother tongue which is different from my own lahjeh (accent), it is always possible to understand them quite well.	Agree	19.2%	28.6%	39.3%	48.3%	50.0%
	Somewhat agree	42.3%	51.4%	39.3%	27.6%	33.3%
	Somewhat disagree	7.7%	2.9%	3.3%	6.9%	16.7%
	Disagree	15.4%	10.0%	4.9%	6.9%	0.0%
	Don't know	15.4%	7.1%	13.1%	10.3%	0.0%
C3.4. If someone speaks to me in a guyesh (dialect) of my mother tongue which is different from my own guyesh (dialect), it is always possible to understand them quite well.	Agree	15.4%	12.9%	14.8%	27.6%	50.0%
	Somewhat agree	26.9%	32.9%	39.3%	20.7%	33.3%
	Somewhat disagree	3.8%	8.6%	13.1%	24.1%	16.7%
	Disagree	23.1%	15.7%	8.2%	13.8%	0.0%
	Don't know	30.8%	30.0%	24.6%	13.8%	0.0%

5.1.3.6. C3-MI (Mutual Intelligibility): Gender

Male participants slightly agreed more with C3 statements compared to female participants. An average of 78% of male and 72% of female participants agreed with C3.1 and C3.2. The majority of male and female participants (85.5% and 71.5% respectively) agreed with C3.3, whereas close to 50% of the participants in each group agreed with C3.4 (see Table 63).

Table 63. Male and female perspectives in C3-MI (Mutual Intelligibility)

		Gender	
		Male	Female
C3.1. If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate <i>zabān</i> (languages).	Agree	49.3%	30.9%
	Somewhat agree	24.6%	38.2%
	Somewhat disagree	2.9%	5.7%
	Disagree	15.9%	17.1%
	Don't know	7.2%	8.1%
C3.2. I cannot understand a language variety if I am not familiar with its words.	Agree	55.1%	35.8%
	Somewhat agree	27.5%	39.0%
	Somewhat disagree	2.9%	4.1%
	Disagree	4.3%	17.1%
	Don't know	10.1%	4.1%
C3.3. If someone speaks to me in a <i>lahjeh</i> (accent) of my mother tongue which is different from my own <i>lahjeh</i> (accent), it is always possible to understand them quite well.	Agree	46.4%	27.6%
	Somewhat agree	39.1%	43.9%
	Somewhat disagree	4.3%	4.9%
	Disagree	4.3%	10.6%
	Don't know	5.8%	13.0%
C3.4. If someone speaks to me in a <i>guyesh</i> (dialect) of my mother tongue which is different from my own <i>guyesh</i> (dialect), it is always possible to understand them quite well.	Agree	18.8%	16.3%
	Somewhat agree	30.4%	33.3%
	Somewhat disagree	10.1%	13.0%
	Disagree	15.9%	12.2%
	Don't know	24.6%	25.2%

5.1.4. C4-SW (Shared Words)

5.1.4.1. Overall results

C4 statements seek to identify whether the extent of shared words among language varieties impacts folk perspectives in identifying language varieties. Participants were requested to identify varieties which share less than 50% and those which share 80% or more of their words as *zabān*, *guyesh* or *lahjeh*. Almost 60% of the participants agreed with C4.1, that two language varieties which share less than 50% of their words are generally separate *zabān*. The majority (80%) of participants agreed with C4.2, that two language varieties which share more than 80% of their words are generally *lahjeh* of the same *zabān* and less than 10% opposed it. Lastly, 40% of the participants agreed with C4.3 and 40% of them disagreed with it (see Table 64). As data suggests,

the majority of participants were in agreement in identifying *lahjeh* of the same *zabān* as well as *zabān* in terms of the extent of shared words. In identifying *guyesh*, however, the participants expressed varied perspectives.

Table 64. C4-SW (Shared Words): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C4.1. If two language varieties share less than 50% of their words, they are generally separate <i>zabān</i> (languages).	21.9%	38.5%	6.8%	15.1%	17.7%
C4.2. If two language varieties share more than 80% of their words, they are generally <i>lahjeh</i> (accents) of the same <i>zabān</i> (language).	39.1%	40.6%	3.6%	6.3%	10.4%
C4.3. If two language varieties share more than 80% of their words, they are generally <i>guyesh</i> (dialects) of the same <i>zabān</i> (language).	15.6%	25.5%	9.9%	32.8%	16.1%

5.1.4.2. C4-SW (Shared Words): Mother tongue

An average of 64.5% of Farsi, Turkic, Mazandarani as well as 54% of Kurdish speakers agreed with C4.1. Additionally, nearly 50% of Mazandarani speakers disagreed with C4.2, whereas only 7% of all the other speakers disagreed with it. Regarding C4.3, a little over 50% of Kurdish participants as well as nearly 37% of Farsi, Turkic and Mazandarani speakers disagreed with it (see Table 65).

Table 65. Farsi, Kurdish, Turkic, and Mazandarani speakers' perspectives in C4-SW (Shared Words)

		Mother tongue			
		Farsi	Kurdish	Turkic	Mazandarani
C4.1. If two language varieties share less than 50% of their words, they are generally separate <i>zabān</i> (languages).	Agree	21.1%	20.0%	23.6%	25.0%
	Somewhat agree	42.1%	34.3%	40.3%	41.7%
	Somewhat disagree	2.6%	5.7%	8.3%	16.7%
	Disagree	10.5%	21.4%	11.1%	16.7%
	Don't know	23.7%	18.6%	16.7%	0.0%
C4.2. If two language varieties share more than 80% of their words, they are generally <i>lahjeh</i> (accents) of the same <i>zabān</i> (language).	Agree	36.8%	38.6%	44.4%	16.7%
	Somewhat agree	36.8%	44.3%	40.3%	33.3%
	Somewhat disagree	0.0%	2.9%	2.8%	25.0%
	Disagree	5.3%	5.7%	4.2%	25.0%
	Don't know	21.1%	8.6%	8.3%	0.0%
C4.3. If two language varieties share more than 80% of their words, they are generally <i>guyesh</i> (dialects) of the same <i>zabān</i> (language).	Agree	15.8%	10.0%	20.8%	16.7%
	Somewhat agree	18.4%	27.1%	29.2%	16.7%
	Somewhat disagree	5.3%	10.0%	9.7%	25.0%
	Disagree	36.8%	41.4%	26.4%	8.3%
	Don't know	23.7%	11.4%	13.9%	33.3%

5.1.4.3. C4-SW (Shared Words): Formal linguistic knowledge

Nearly 50% of the participants who had taken courses in linguistics agreed with C4.1. Among the participants without formal linguistic knowledge, nearly 62% agreed with it. Both groups reached a similar level of agreement in identifying *lahjeh* and *guyesh* in C4.2 and C4.3. That said, approximately 77% of both groups agreed with C4.2 and 40% of them agreed with C4.3 (see Table 66).

Table 66. Participants' perspectives with respect to formal linguistic knowledge in C4-SW (Shared Words)

		Formal linguistic knowledge	
		No	Yes
C4.1. If two language varieties share less than 50% of their words, they are generally separate zabān (languages).	Agree	22.1%	21.1%
	Somewhat agree	40.3%	31.6%
	Somewhat disagree	8.4%	0.0%
	Disagree	14.9%	15.8%
	Don't know	14.3%	31.6%
C4.2. If two language varieties share more than 80% of their words, they are generally lahjah (accents) of the same zabān (language).	Agree	37.7%	44.7%
	Somewhat agree	43.5%	28.9%
	Somewhat disagree	4.5%	0.0%
	Disagree	6.5%	5.3%
	Don't know	7.8%	21.1%
C4.3. If two language varieties share more than 80% of their words, they are generally guyesh (dialects) of the same zabān (language).	Agree	14.9%	18.4%
	Somewhat agree	26.6%	21.1%
	Somewhat disagree	11.7%	2.6%
	Disagree	32.5%	34.2%
	Don't know	14.3%	23.7%

5.1.4.4. C4-SW (Shared Words): Education

Almost 60% of the participants in each educational category agreed with C4.1 (see Table 67). Although most college, BA and MA/PhD participants agreed with C4.2, graduate participants agreed most (86%) with this statement. Results showed that except for the BA participants, the number of those who opposed C4.3 (47.5%) was higher than of those who agreed (37.5%) with it.

Table 67. College, BA and MA/PhD participants' perspectives regarding C4-SW (Shared Words)

		Education		
		College	BA	MA/PhD
C4.1. If two language varieties share less than 50% of their words, they are generally separate zabān (languages).	Agree	31.8%	13.3%	21.6%
	Somewhat agree	27.3%	46.7%	41.2%
	Somewhat disagree	6.1%	8.0%	5.9%
	Disagree	18.2%	14.7%	11.8%
	Don't know	16.7%	17.3%	19.6%
C4.2. If two language varieties share more than 80% of their words, they are generally lahjeh (accents) of the same zabān (language).	Agree	50.0%	29.3%	39.2%
	Somewhat agree	30.3%	45.3%	47.1%
	Somewhat disagree	1.5%	8.0%	0.0%
	Disagree	9.1%	6.7%	2.0%
	Don't know	9.1%	10.7%	11.8%
C4.3. If two language varieties share more than 80% of their words, they are generally guyesh (dialects) of the same zabān (language).	Agree	16.7%	14.7%	15.7%
	Somewhat agree	19.7%	32.0%	23.5%
	Somewhat disagree	13.6%	9.3%	5.9%
	Disagree	36.4%	25.3%	39.2%
	Don't know	13.6%	18.7%	15.7%

5.1.4.5. C4-SW (Shared Words): Age

Nearly 66% of the participants in the age categories of '20-30' and 'over 50' and nearly 54% of the participants in all other age categories agreed with C4.1 (see Table 68). An average of 80% of the participants under 50 years of age and 67% of 'over 50' agreed with C4.2. Almost 33% of 'over 50' year-old participants disagreed with C4.2 while this number was as low as nearly 10% in other age categories. Participants' decisions were not as unanimous about C4.3 as was the case for C4.2 (see Table 68).

Table 68. Participants' perspectives with respect to age in C4-SW (Shared Words)

		Age				
		Under 20	20-30	30-40	40-50	Over 50
C4.1. If two language varieties share less than 50% of their words, they are generally separate zabān (languages).	Agree	15.4%	20.0%	24.6%	24.1%	33.3%
	Somewhat agree	38.5%	50.0%	31.1%	27.6%	33.3%
	Somewhat disagree	11.5%	2.9%	3.3%	17.2%	16.7%
	Disagree	26.9%	7.1%	19.7%	17.2%	0.0%
	Don't know	7.7%	20.0%	21.3%	13.8%	16.7%
C4.2. If two language varieties share more than 80% of their words, they are generally lahjah (accents) of the same zabān (language).	Agree	30.8%	34.3%	41.0%	55.2%	33.3%
	Somewhat agree	46.2%	42.9%	41.0%	31.0%	33.3%
	Somewhat disagree	11.5%	2.9%	1.6%	0.0%	16.7%
	Disagree	3.8%	7.1%	6.6%	3.4%	16.7%
	Don't know	7.7%	12.9%	9.8%	10.3%	0.0%
C4.3. If two language varieties share more than 80% of their words, they are generally guyesh (dialects) of the same zabān (language).	Agree	19.2%	15.7%	9.8%	17.2%	50.0%
	Somewhat agree	34.6%	21.4%	18.0%	37.9%	50.0%
	Somewhat disagree	15.4%	7.1%	6.6%	20.7%	0.0%
	Disagree	26.9%	34.3%	45.9%	13.8%	0.0%
	Don't know	3.8%	21.4%	19.7%	10.3%	0.0%

5.1.4.6. C4-SW (Shared Words): Gender

Almost 66% of male and 58.5% of female participants agreed with C4.1 (see Table 69). Most of the male participants agreed with C4.2 and only 29% of them agreed with C4.3. Among female participants, 78% agreed with C4.2 and nearly 50% agreed with C4.3.

Table 69. Male and female perspectives in C4-SW (Shared Words)

		Gender	
		Male	Female
C4.1. If two language varieties share less than 50% of their words, they are generally separate zabān (languages).	Agree	29.0%	17.9%
	Somewhat agree	34.8%	40.7%
	Somewhat disagree	7.2%	6.5%
	Disagree	7.2%	19.5%
	Don't know	21.7%	15.4%
C4.2. If two language varieties share more than 80% of their words, they are generally lahjah (accents) of the same zabān (language).	Agree	44.9%	35.8%
	Somewhat agree	37.7%	42.3%
	Somewhat disagree	1.4%	4.9%
	Disagree	2.9%	8.1%
	Don't know	13.0%	8.9%
C4.3. If two language varieties share more than 80% of their words, they are generally guyesh (dialects) of the same zabān (language).	Agree	13.0%	17.1%
	Somewhat agree	15.9%	30.9%
	Somewhat disagree	8.7%	10.6%
	Disagree	43.5%	26.8%
	Don't know	18.8%	14.6%

5.1.5. C5-WS (Writing System)

5.1.5.1. Overall results

C5 statements seek language identification with respect to a writing system (see Table 70). Nearly 41% of the participants agreed and 21% disagreed with C5.1, that if a language variety has a standardized writing system, it may not be a *zabān*. These perspectives could reflect folk attitude towards the official language of the country, due to intensive interactions with this language and their lack of desire to be unified under one writing system. As scholars (e.g., see Cahill & Karan, 2008; Karan, 2006; Robinson & Gadellii, 2003) have pointed out, a writing system does not solely reflect speech communities' desires in capturing linguistic features of their language varieties in written form and how these features may be different from those of an official language in the country. Sociocultural tendencies of speech communities to differentiate themselves from other communities may range from a writing system with a completely different script to one with slightly different orthographic conventions (e.g., see Priestly, 1992; Sebba, 1998).

Table 70. C5-WS (Writing System): Likert-scale responses

	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
C5.1. If a language variety has a standardized writing system, it may not be a <i>zabān</i> (language).	17.2%	24.0%	6.3%	14.1%	38.5%
C5.2. <i>Guyesh</i> (dialects) do not have standardized writing systems.	17.7%	22.9%	12.0%	24.5%	22.9%
C5.3. To be considered a <i>zabān</i> (language), a language variety needs to have a standardized writing system.	38.0%	43.8%	4.2%	4.7%	9.4%

Approximately 41% of the participants agreed with C5.2, that a *guyesh* does not have a writing system while 36.5% of the participants disagreed with it. Further, 82% agreed with C5.3, that to be considered a *zabān*, a variety must have a standardized writing system. Table 70 shows that the highest level of agreement was reached regarding C5.3.

Such complications, once again, point to the fact that singular parameters cannot determine whether a language variety is a *zabān*, *guyesh* or *lahjeh*. As the literature has pointed out, every speech community identifies a circle of intertwined parameters which collectively help individuals identify language varieties. One singled-out parameter, although influential, may not be sufficient (Saville-Troike, 2003). Additionally, the ‘level of divergence’ sufficient to distinguish different types of language varieties and apply linguistic terms such as *zabān*, *guyesh* or *lahjeh* is not necessarily based on the same set of criteria (see Eberhard et al., 2021). By the same token, priority given to influential parameters are not equally weighted in different speech communities (Wolfram, 2017).

Further, while a writing system was considered as a necessary parameter to identify a language variety as *zabān*, for *guyesh*, such definitive results were not apparent. That is, whether or not a *guyesh* has a writing system was yet another controversial matter among the participants. Lack of agreement in this regard is also reflected in scholarly discussions. Some linguists argue that except for the standard variety of a country, the rest of the regional and social varieties should concentrate on speech only. This perspective, however, does not consider the fact that some of the language varieties already have an established writing tradition. For example, as one of the minority languages spoken in Iran, Kurdish has two well-established writing systems that have been in use for nearly a century (see Hassanpour, 1992). Eira (1998, p. 176) further adds that “whether or not one can propose principles for a linguistically optimal writing system, it does not at all follow that linguistic efficacy is the only or the most significant factor in the creation of orthography, defined as the accepted standard for writer/readers of the language”.

5.1.5.2. C5-WS (Writing System): Mother tongue

An average of 37% of each of Farsi and Turkic, and 48% of each of Kurdish and Mazandarani speakers agreed with C5.1 (see Table 71). Almost 37.5% in each language group chose ‘don’t know’ in C5.1. Regarding C5.2, nearly 58% of Farsi, 27% of Kurdish, 32% of Turkic, and 50% of Mazandarani speakers disagreed with it. Lastly, most of the participants, regardless of their mother tongue, agreed with C5.3.

Table 71. Farsi, Kurdish, Turkic, and Mazandarani speakers’ perspectives in C5-WS (Writing System)

		Mother tongue			
		Farsi	Kurdish	Turkic	Mazandarani
C5.1. If a language variety has a standardized writing system, it may not be a <i>zabān</i> (language).	Agree	18.4%	20.0%	13.9%	16.7%
	Somewhat agree	18.4%	25.7%	23.6%	33.3%
	Somewhat disagree	5.3%	4.3%	9.7%	0.0%
	Disagree	15.8%	15.7%	11.1%	16.7%
	Don't know	42.1%	34.3%	41.7%	33.3%
C5.2. <i>Guyesh</i> (dialects) do not have standardized writing systems.	Agree	7.9%	21.4%	18.1%	25.0%
	Somewhat agree	18.4%	18.6%	29.2%	25.0%
	Somewhat disagree	18.4%	10.0%	11.1%	8.3%
	Disagree	39.5%	17.1%	20.8%	41.7%
	Don't know	15.8%	32.9%	20.8%	0.0%
C5.3. To be considered a <i>zabān</i> (language), a language variety needs to have a standardized writing system.	Agree	47.4%	34.3%	34.7%	50.0%
	Somewhat agree	44.7%	45.7%	43.1%	33.3%
	Somewhat disagree	2.6%	4.3%	5.6%	0.0%
	Disagree	2.6%	5.7%	4.2%	8.3%
	Don't know	2.6%	10.0%	12.5%	8.3%

5.1.5.3. C5-WS (Writing System): Formal linguistics knowledge

More than 50% of the participants who had taken linguistics courses chose ‘don’t know’, 34% agreed with C5.1. Among those without formal linguistic knowledge, 43% agreed with C5.1 (see Table 72). An average of 39% of the participants, both with and without formal linguistic knowledge, agreed with C5.2. Results further indicated that nearly 75% of the participants who

had taken linguistics courses agreed with C5.3. Among those with no formal linguistic knowledge, nearly 85% of them agreed with C5.3.

Table 72. Participants’ perspectives with respect to formal linguistic knowledge in C5-WS (Writing System)

		Formal linguistic knowledge	
		No	Yes
C5.1. If a language variety has a standardized writing system, it may not be a zabān (language).	Agree	16.9%	18.4%
	Somewhat agree	26.0%	15.8%
	Somewhat disagree	6.5%	5.3%
	Disagree	16.2%	5.3%
	Don't know	34.4%	55.3%
C5.2. Guyesh (dialects) do not have standardized writing systems.	Agree	16.9%	21.1%
	Somewhat agree	24.7%	15.8%
	Somewhat disagree	13.6%	5.3%
	Disagree	23.4%	28.9%
	Don't know	21.4%	28.9%
C5.3. To be considered a zabān (language), a language variety needs to have a standardized writing system.	Agree	38.3%	36.8%
	Somewhat agree	45.5%	36.8%
	Somewhat disagree	4.5%	2.6%
	Disagree	5.8%	0.0%
	Don't know	5.8%	23.7%

5.1.5.4. C5-WS (Writing System): Education

In C5, college, BA and MA/PhD participants’ perspectives were not significantly different (see Table 73). An average of 41% of the participants from all educational background agreed with C5.1. The population of BA and MA/PhD participants who agreed with C5.2 (45% and 40%, respectively) is almost 10% more than those who disagreed with it. Among college degree holders, 44% agreed and 36% disagreed with C5.2. Lastly, most college, BA and MA/PhD participants agreed with C5.3.

Table 73. College, BA and MA/PhD participants' perspectives regarding C5-WS (Writing System)

		Education		
		College	BA	MA/PhD
C5.1. If a language variety has a standardized writing system, it may not be a zabān (language).	Agree	19.7%	14.7%	17.6%
	Somewhat agree	24.2%	25.3%	21.6%
	Somewhat disagree	7.6%	5.3%	5.9%
	Disagree	15.2%	14.7%	11.8%
	Don't know	33.3%	40.0%	43.1%
C5.2. Guyesh (dialects) do not have standardized writing systems.	Agree	18.2%	17.3%	17.6%
	Somewhat agree	18.2%	28.0%	21.6%
	Somewhat disagree	15.2%	10.7%	9.8%
	Disagree	28.8%	25.3%	17.6%
	Don't know	19.7%	18.7%	33.3%
C5.3. To be considered a zabān (language), a language variety needs to have a standardized writing system.	Agree	36.4%	38.7%	39.2%
	Somewhat agree	48.5%	45.3%	35.3%
	Somewhat disagree	6.1%	4.0%	2.0%
	Disagree	4.5%	5.3%	3.9%
	Don't know	4.5%	6.7%	19.6%

5.1.5.5. C5-WS (Writing System): Age

Findings showed that with age, the number of those who agreed with C5.1 increased. Regarding this, an average of 37% of 'under 20', '20-30', '30-40', nearly 50% of '40-50', and 67% of 'over 50' year-old participants agreed with C5.1. Many of the participants in the first three age categories (see Table 74) chose 'don't know' regarding C5.1, whereas the participants in '40-50' age category disagreed most with it. An estimate of 36% of the participants in the age categories of '20-30' and '40-50' as well as 44% of 'under 20' and '40-50' year-old participants agreed with C5.2. Except for the participants older than 40, 34% of all other age categories disagreed with C5.2. Finally, approximately more than 80% of the participants across all age categories agreed with C5.3.

Table 74. Participants' perspectives with respect to age in C5-WS (Writing System)

		Age				
		Under 20	20-30	30-40	40-50	Over 50
C5.1. If a language variety has a standardized writing system, it may not be a zabān (language).	Agree	11.5%	10.0%	24.6%	20.7%	33.3%
	Somewhat agree	23.1%	25.7%	16.4%	34.5%	33.3%
	Somewhat disagree	11.5%	5.7%	0.0%	17.2%	0.0%
	Disagree	7.7%	20.0%	11.5%	10.3%	16.7%
	Don't know	46.2%	38.6%	47.5%	17.2%	16.7%
C5.2. Guyesh (dialects) do not have standardized writing systems.	Agree	11.5%	11.4%	21.3%	24.1%	50.0%
	Somewhat agree	34.6%	25.7%	19.7%	10.3%	33.3%
	Somewhat disagree	7.7%	8.6%	8.2%	34.5%	0.0%
	Disagree	26.9%	25.7%	24.6%	20.7%	16.7%
	Don't know	19.2%	28.6%	26.2%	10.3%	0.0%
C5.3. To be considered a zabān (language), a language variety needs to have a standardized writing system.	Agree	46.2%	34.3%	39.3%	27.6%	83.3%
	Somewhat agree	34.6%	48.6%	42.6%	51.7%	0.0%
	Somewhat disagree	7.7%	1.4%	1.6%	13.8%	0.0%
	Disagree	7.7%	1.4%	6.6%	3.4%	16.7%
	Don't know	3.8%	14.3%	9.8%	3.4%	0.0%

5.1.5.6. C5-WS (Writing System): Gender

In terms of gender responses to questions about writing systems, 46% of male and 38% of female participants agreed with C5.1 and an average of 38% of both male and female participants chose 'don't know'. An equal percentage of male participants (37.5%) disagreed and agreed with C5.2. Among female participants, however, the population of those who agreed with C5.2 was slightly higher than those who disagreed with it. Lastly, although most of the participants agreed with C5.3, the number of male participants was slightly higher than females (see Table 75).

Table 75. Male and female perspectives in C5-WS (Writing System)

		Gender	
		Male	Female
C5.1. If a language variety has a standardized writing system, it may not be a <i>zabān</i> (language).	Agree	17.4%	17.1%
	Somewhat agree	29.0%	21.1%
	Somewhat disagree	4.3%	7.3%
	Disagree	11.6%	15.4%
	Don't know	37.7%	39.0%
C5.2. <i>Guyesh</i> (dialects) do not have standardized writing systems.	Agree	15.9%	18.7%
	Somewhat agree	20.3%	24.4%
	Somewhat disagree	11.6%	12.2%
	Disagree	27.5%	22.8%
	Don't know	24.6%	22.0%
C5.3. To be considered a <i>zabān</i> (language), a language variety needs to have a standardized writing system.	Agree	37.7%	38.2%
	Somewhat agree	47.8%	41.5%
	Somewhat disagree	4.3%	4.1%
	Disagree	1.4%	6.5%
	Don't know	8.7%	9.8%

Overall, analysis of C1-GN findings revealed that more than half of the participants disagreed with the statements which investigate the relevance of geography and number of speakers in language identification whereas nearly one-third of these participants agreed with them. These perspectives, which align with the findings of both the interviews and pilot iterations, underline the heterogeneity of perspectives established around the meaningful contribution of these parameters. C2-ST results determined *guyesh* and *lahjeh* as language varieties which, according to the government, have no official status. Regarding their use, however, folk perspectives varied. That is, while *guyesh* was generally regarded as having no use in the education system, this was not the case for *lahjeh*. The results from C3-MI indicated that the majority of participants identified both *lahjeh* and *zabān* in terms of the extent of mutual intelligibility between speakers of these varieties. *Guyesh*, however, turned out to be a more challenging type of variety to define and differentiate compared with *lahjeh* and *zabān*. Not only in C3-MI, but also in C4-SW and C5-WS identifying *guyesh* in terms of the extent of shared words and writing system was

difficult. That is, the participants did not reach an agreement regarding the extent to which two *guyesh* share words and whether or not, similar to *zabān*, *guyesh* has a standardized writing system.

5.1.6. Multiple regression analysis

To investigate whether all of the participants' demographic categories contributed significant unique predictive power to the regression when considered in combination with other predictors, I ran hierarchical regression analysis. Using this type of regression analysis, I investigated the influence of a model with all the multiple-category predictors as well as the amount of the contribution of every independent predictor (variable) introduced into the model. To run regression analysis, Likert-scale responses were treated as interval. The summed score of Likert-scale data was adopted as dependent and the participants' demographics as independent variables.

In running regression, researchers are interested in explaining the most variability in the dependent variable with the fewest possible predictors (parsimony). Since the independent variables were categorical (qualitative), to run hierarchical regression analysis, first, I had to incorporate categorical predictors through the use of dummy variables. Dummy coding is a method of converting the grouping information of categorical variables into a new set of (dichotomous) variables that can be used as predictors in a regression model. Each of these variables indicates the presence or absence of certain attributes. A dummy variable assumes value 1 if the attribute is present and value 0 if it is absent. For instance, in this study, the dummy coding of a multiple-category variable such as education led to three dummy variables (i.e., college, BA, MA/PhD). When one of these three variables assumed the value 1, the rest of them assumed the value 0. For

example, for the college graduate participants, 'college' assumed the value 1. For these participants, the values for 'BA' and 'MA/PhD' were 0. Similarly, for those participants who identified their level of education as 'BA', this dummy variable assumed the value 1 and all other educational categories assumed the value 0. I applied this coding system to all of the multiple-category variables of this study. That said, not all variables needed to be dummy coded. Variables such as gender (coded 0=Male, 1=Female) and 'formal linguistic knowledge' (coded 0=No, 1=Yes) in the dataset already contained this coding. Thus, I did not recode them.

Regression analysis can incorporate grouping variables if they have been appropriately recoded into a new set of dummy variables. This new set equals the number of groups on the original variable minus 1 (i.e., $K-1$). The variable that is omitted becomes the comparison group against which the other variables are compared. To properly interpret the regression results for the variables which were dummy coded, I established a baseline (or reference category) against which all other groups were compared. This way, the regression coefficients for each dummy variable were equal to the difference in conditional means on the dependent variable between a specific group and the reference category. This is to say that dummy variable coefficients are interpreted as the change in the predicted value of the dependent variable. For example, to examine whether the participants' education statistically significantly contributed to the variation in their overall scores, from among the dummy variables for this predictor, I established 'college' as the reference category. That is, the intercept in the model summary was the mean for the college graduates. This way, other dummy variables in the education category (i.e., BA and MA/PhD) were compared

against college as the reference category. This was the procedure for all of the categorical variables which were coded using dummy coding.

It is also noteworthy that any group can be treated as the reference category, with all the dummy variables capturing differences between the conditional mean for the reference category and the remaining groups. In this study, considering that the purpose was to investigate the influence of all the independent variables, the intercept (constant) was the conditional mean for the participants who (a) fell into the ‘college’ education group, (b) with no formal linguistic knowledge, (c) identified with Farsi as mother tongue, (d) fell into the ‘under 20’ age group, and (e) identified as male.

In the model summary table (see Table 76), R is a measure of the prediction quality of the dependent variable whereas R square provides a prediction of how much of the dependent variable is explained by the independent variables. As the model summary table shows, in every model, a new multiple-category independent variable is introduced. For example, model 1 has only one multiple-category independent variable (i.e., education) whereas model two has two independent variables. R square value in model 1 equals .37. This means that the proposed regression model with one multiple-category variable (i.e., education) explains 37% of the variation in the participants’ overall scores.

Table 76. Regression model summary

Model	R	RSquare	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					RSquare Change	F Change	df1	df2	Sig F Change
1	.61 ^a	.37	.36	6.95	.37	55.02	2	189	.000
2	.70 ^b	.49	.48	6.29	.12	42.85	1	188	.000
3	.73 ^c	.53	.52	6.04	.05	6.15	3	185	.001
4	.74 ^d	.54	.52	6.05	.01	.84	4	181	.503
5	.74 ^e	.54	.51	6.07	.00	.18	1	180	.668

a. Predictors: (Constant), MA/PhD, BA

b. Predictors: (Constant), MA/PhD, BA, Formal LING knowledge

c. Predictors: (Constant), MA/PhD, BA, Formal LING knowledge , Turkic, Mazandarani, Kurdish

d. Predictors: (Constant), MA/PhD, BA, Formal LING knowledge , Turkic, Mazandarani, Kurdish, Over 50, 30-40, 40-50, 20-30

e. Predictors: (Constant), MA/PhD, BA, Formal LING knowledge , Turkic, Mazandarani, Kurdish, Over 50, 30-40, 40-50, 20-30, Gender

Similarly, model 2 is significant with two multiple-category variables, namely: education and formal linguistic knowledge. As the ‘R square change’ column in model 2 indicates, the introduction of ‘formal linguistic knowledge’ contributed an additional 12% to the R square. That is, 49% of the variation in the participants’ overall score is explained by ‘level of education’ and ‘formal linguistic knowledge’. In model 3, the introduction of a third independent variable (i.e., mother tongue) significantly increased the R square to .53. However, the introduction of ‘age’ in model 4 and ‘gender’ in model 5 did not statistically significantly contribute to the amount of variation in the overall scores. Thus, while the participants’ education, formal linguistic knowledge and mother tongue influenced their overall scores, the predictive power of age and gender was not significant (see Table 76).

The analysis of variance is used to test the statistical significance of the R-square value in the model summary table. The null hypothesis is that the population R-square is zero. ANOVA

results (see model 5 in Table 77) indicated statistical significance [$F(11,180)=19.28, p<.001$]. This suggests that the independent variables statistically significantly predicted the dependent variable. Although not all components of this regression model significantly predicted the variation in the participants' overall scores, not all of the population regression coefficients were equal to 0. This means that the null hypothesis which suggested that all population regression coefficients associated with the model were equal to 0 can be rejected (see Lomax & Hahs-Vaughn, 2012).

Table 77. Model fitness via ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5311.28	2	2655.64	55.02	.000 ^b
	Residual	9121.72	189	48.26		
	Total	14433.00	191			
2	Regression	7004.44	3	2334.81	59.09	.000 ^c
	Residual	7428.56	188	39.51		
	Total	14433.00	191			
3	Regression	7677.75	6	1279.62	35.04	.000 ^d
	Residual	6755.25	185	36.51		
	Total	14433.00	191			
4	Regression	7800.51	10	780.05	21.29	.000 ^e
	Residual	6632.49	181	36.64		
	Total	14433.00	191			
5	Regression	7807.30	11	709.75	19.28	.000 ^f
	Residual	6625.70	180	36.81		
	Total	14433.00	191			

a. Dependent Variable: Sum of responses

b. Predictors: (Constant), MA/PhD, BA

c. Predictors: (Constant), MA/PhD, BA, Formal LING knowledge

d. Predictors: (Constant), MA/PhD, BA, Formal LING knowledge, Turkic, Mazandarani, Kurdish

e. Predictors: (Constant), MA/PhD, BA, Formal LING knowledge, Turkic, Mazandarani, Kurdish, Over 50, 30-40, 40-50, 20-30

f. Predictors: (Constant), MA/PhD, BA, Formal LING knowledge, Turkic, Mazandarani, Kurdish, Over 50, 30-40, 40-50, 20-30, Gender

To understand how individual dummy variables impacted the participants' overall scores, I examined the coefficients of these variables in model 5 (see Table 78). In a coefficients table, the 'constant' is the intercept in the model. The intercept is the predicted value on Y (dependent variable) when the predictors are all 0. The slope for the predictor (coefficient value) indicates the predicted change in Y for one unit increase on X (independent variable). This is essentially the difference in means between values coded as 0 and values coded as 1 (e.g., coded 0=Male, 1=Female). As a result, the slope is interpreted as indicating that, on average, a value coded as 1 scores Z points higher (+) or lower (-) than a value coded as 0. Then the mean for values coded as 1 is simply computed as the sum of the intercept and the slope.

Table 78. Coefficients, correlations and collinearity statistics in model 5 with five variables

Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error				Zero-order	Partial	Part	Tolerance	VIF
5	(Constant)	64.15	2.07		30.99	.000					
	BA	10.91	1.38	.61	7.93	.000	.53	.51	.40	.43	2.35
	MA/PhD	8.93	1.33	.46	6.71	.000	.01	.45	.34	.55	1.80
	Formal LING knowledge	-7.82	1.21	-.36	-6.47	.000	-.27	-.43	-.33	.83	1.21
	Mazandarani	1.90	2.10	.05	.90	.368	.22	.07	.05	.74	1.35
	Turkic	2.89	1.24	.16	2.34	.020	.25	.17	.12	.53	1.87
	Kurdish	-.88	1.28	-.05	-.69	.492	-.33	-.05	-.03	.51	1.97
	20-30	-1.85	1.54	-.10	-1.20	.233	.13	-.09	-.06	.35	2.88
	30-40	-2.94	1.74	-.16	-1.69	.093	-.26	-.12	-.09	.29	3.42
	40-50	-2.48	1.97	-.10	-1.26	.209	-.23	-.09	-.06	.39	2.59
	Over 50	.34	2.79	.01	.12	.904	.11	.01	.01	.81	1.23
	Gender	-.43	1.00	-.02	-.43	.668	.20	-.03	-.02	.83	1.21

a. Dependent Variable: Sum of responses

In the unstandardized coefficients table (see Table 78), the constant=64.15. As mentioned above, this means that the predicted value on the dependent variable (i.e., overall score in the language identification questionnaire) is 64.15 when all the predictors are zero. For the first

multiple-category predictor (i.e., education), the predicted overall score of a participant who identified as a BA graduate (i.e., coded 1=BA and coded 0=all else) is significantly higher (slope=10.91) than the predicted overall score of a participant in reference or base group (i.e., college group), conditioned on the other predictors in the model. This condition means that one needs to consider the influence of other predictors including formal linguistic knowledge, mother tongue, age, and gender, since none of these predictors are zero, though their influence may be insignificant (e.g., see gender coefficient, $B = -.43$).

In a similar way, MA/PhD holders report an overall score which is higher than college degree holders by 8.93 points. This means that the average scores on the questionnaire for college, BA and MA/PhD holders are 64, 75 and 73, conditioned on the other variables. Similar to education, on average, the overall score of the participants who took courses in linguistics was lower ($[64.15] + [-7.82] = 56.33$) compared with those without formal linguistic knowledge. With respect to mother tongue, regression results revealed that only Turkic speakers' sum of responses were, on average, statistically significantly higher ($B = 2.89$) than the overall score of the reference group (i.e., Farsi speakers). The coefficients for Mazandarani ($B = 1.90$) and Kurdish ($B = -.88$) participants show that their overall scores were not statistically significantly different from that of a Farsi speaker (see Table 78).

Further, the unstandardized coefficients of age categories show that the overall scores of the participants from different age categories were not statistically significantly different. Regarding this, the coefficients indicate that the participants from 20 to 50 years of age scored

lower compared with those in the ‘under 20’ age category. Further, the participants who were over 50 years of age scored slightly higher, conditioned on the other predictors.

The unstandardized coefficient for gender ($B=-.43$) indicates that the average female score on the language identification questionnaire was lower, by less than one point, compared with the average male score (see Table 78). As the significance column reveals, the difference between the perspectives of male and female participants of this study, conditioned on the other predictors in the regression model, is insignificant ($p=.66$).

Considering the unstandardized coefficient B for the independent variables, the participants’ education, formal linguistic knowledge and mother tongue were significant predictors of their linguistic perspectives. However, the changes in the predicted value of the participants’ sum of Likert-scale responses on the language identification questionnaire due to gender and age were not statistically significant. In an attempt to create a parsimonious model, the respective variables are considered excessive. This is because, without these two variables, the R square does not significantly change (see the R square for model 3 compared with models 4 and 5 in Table 76). Thus, the best model which can explain the variation in the way folk identify language varieties is one with three independent variables.

The correlation section of the output (see Table 78) elaborates on the correlations between the independent variables and the dependent variable. In particular, the ‘partial correlations’ column points to the correlation between the participants’ overall scores and each category of the multiple-category demographic variables while controlling the remaining categories. The significance of semi-partial correlations (Part column) is that the square of the semi-partial

correlation (r_{sp}^2) reflects the proportion of variation in the dependent variable that is uniquely accounted for by a given independent variable (Tabachnick & Fidell, 2013). This, in turn, ranks the independent variables in terms of their relative contributions to the regression model. The squared semi-partial values show that ‘education’ had the strongest predictive relationship to the participants’ linguistic perspectives, followed by ‘formal linguistic knowledge’ and ‘mother tongue’.

A fundamental assumption in multiple regression analysis is the absence of multicollinearity among the independent variables in a regression model (Lomax & Hahs-Vaughn, 2012; Pituch & Stevens, 2016). To test multicollinearity, two indices of ‘Tolerance’ and ‘VIF’ should be analyzed in the regression output (see collinearity statistics in Table 78). Tolerance values $< .10$ are indicative of the presence of multicollinearity and $VIF > 10$ can be considered indicative of the presence of more severe multicollinearity (Lomax & Hahs-Vaughn, 2012). In the current demonstration, tolerance values are greater than $.10$. and all VIF values are well below 10, ranging between Gender=1.21 and age (30-40)=3.42.

In light of the findings, hierarchical multiple regression results revealed that the participants’ formal linguistic knowledge, education and mother tongue statistically significantly (53%) predicted the variation in their linguistics perspectives, $F(11,180)=19.28$, $p<.001$, $R^2=.53$. All variables except the participants’ gender and age added to the prediction, $p<.05$. Thus, participants’ demographics statistically significantly contributed to the amount of variation in their overall scores in the language identification questionnaire.

6. Discussion and Conclusion

In all areas of the world, a rich history of research in linguistics has yielded invaluable insight into the lexical similarity, and the phonological, morphological and syntactic systems of languages, and has attempted to address the issues of language identification on this basis. Defining and distinguishing between types of language varieties based on their structural characteristics is treated as self-evident by scholars.

Nonetheless, identifying language varieties with respect to linguistic parameters is on its own insufficient for meaningful identification of language varieties. Languages are complex social entities integrated with culture, identity, ethnicity, religion, social norms of communities, and individuals' ideologies in life (Joseph, 2004). Scholars (e.g., see Edwards, 2009; Hymes, 1967, 1974) stress the significance of the social world to the point that language varieties are both components and reflections of social or cultural contexts. As such, what language is cannot be separated from how and why it is perceived, defined and used a particular way.

Moreover, the approaches that folk (non-linguists) adopt to identify language varieties do not always align with linguistic labels used by specialists, but rather are constructs of how they themselves perceive these varieties (Preston, 2011). While scholars, Iranian or Western, have focused on structural and functional properties of languages in Iran and throughout the world, the subdiscipline of folk linguistics has enriched the discourse around language identification by including perspectives of folk as part of systematic research into the topic. Such studies have been carried out in many parts of the world, with a longstanding hub of activity in Europe. However, very few studies on folk linguistics exist for the languages of Iran.

Available literature on language identification comments, to a limited extent, on directionality or strength of correlations among parameters associated with folk linguistic perspectives. It mostly underlines that intertwined parameters are at work in addressing these issues (Hymes, 1974; Preston, 1989, 1999a). Further, addressing these issues may take different paths depending on demographics of participants (Saville-Troike, 2003).

To begin to understand how folk linguistic perspectives were formed, it was clear that the issues of language identification could not be adequately addressed within the traditional confines of either structural or functional approaches to language as the issues of human attitudes and beliefs make the study of language in use “a thoroughly interdisciplinary enterprise” (Coupland 1988, p. 19). For these reasons, my aim became that of articulating an approach to the study of language varieties which would draw on the established paradigms of sociolinguistics, folk linguistics, and perceptual dialectology and by exemplifying the sorts of analyses that such orientations permitted.

Scarcity of tools to address the issues of language identification from people’s perspectives on the one hand and shortcomings inherent in methods such as interviews on the other hand, added to the challenge of addressing these issues. Responding to the insufficiency of studies with a focus on language identification in Iran and informed by linguistic, sociolinguistic and folk linguistic theories, the purpose of this research project was to develop, validate and carry out a language identification questionnaire to collect folk linguistic perspectives which would investigate how folk define and differentiate different types of language varieties in Iran (research question #1) and further identify parameters which, with respect to folk perspectives, significantly contribute to language identification (research question #2). It is noteworthy that, in this questionnaire, I used

Farsi linguistic labels in all administration occasions. These labels, though not fully consistently defined by scholars, included *zabān* (roughly equivalent to the English word for language), *guyesh* (roughly equivalent to the English word for dialect) and *lahjeh* (roughly equivalent to the English word for accent). Also, to identify the relevant constructs in language identification and develop and validate this questionnaire, I followed Hinkin's (1998) framework which guided this research project in eleven steps.

In line with data collected from the pilot iterations, the administration of the questionnaire revealed significant results regarding how folk with different demographic parameters defined and distinguished *zabān*, *guyesh* and *lahjeh*. The results showed that the participants expressed diverse perspectives in the statements related to the relevance of geography and number of speakers (C1-GN) in language identification. Nonetheless, their views in the rest of this questionnaire confirmed the meaningful contribution of status (C2-ST), mutual intelligibility (C3-MI), shared words (C4-SW), and standardized writing system (C5-WS) in identifying language varieties.

In identifying different types of language varieties (research question #1) in terms of the respective parameters (research question #2), the participants defined *guyesh* as a language variety which is neither used in the education system nor has an official status according to the government. Definition and differentiation of *guyesh* in terms of the extent of mutual intelligibility and shared words as well as a writing system was a controversial matter among the participants. *Lahjeh* of the same *zabān* were more often identified as mutually intelligible language varieties with no official status, according to the government. Two *zabān*, on the other hand, were defined as mutually unintelligible language varieties which have writing systems. It can also be inferred

that people are not familiar, to a great extent, with the words of a separate *zabān*, whereas they are familiar with and understand the words of a *lahjeh* which belongs to their mother tongue. The majority of participants identified two varieties which share less than 50% and more than 80% of their words as *zabān* and *lahjeh*, respectively.

It is noteworthy that the relevance of each of the respective parameters was established merely based on the participants' extent of agreement with the related statements in each component and not with respect to these parameters in general. In case future researchers modify the questionnaire and validate it with different statements in any of the components, the results might well be different.

In identifying parameters which significantly contributed to language identification (research question #2), the participants' demographics proved to be influential (see Step eleven). Regression analysis revealed that the participants' education, formal linguistic knowledge and mother tongue predicted 53% of the variation in their linguistic perspectives. Taking these parameters into account individually, though influential, does not represent the same impact. As Saville-Troike (2003) asserts, certain parameters are influential only as part of an intertwined circle of parameters. As such, it is important to value all the suggested parameters as valid and view the list as a cumulative but not an exhaustive one.

With respect to language identification, regression analysis results indicated that the participants' perspectives about types of language varieties did not differ regarding whether they were male or female. This is to say that their biological orientation was not the underlying reason for the way they defined and differentiated types of language varieties. Rather, their gender was a

socially constructed category delimited to a male-female dimension (Saville-Troike, 2003, p. 77) which was not brought off as significant in the way they addressed the issues of language identification. Thus, gender was not underlined as statistically significantly contributing to language identification.

The fact that the predictive power of gender and age turned out to be insignificant in the regression model of this study (see Table 76) only suggests that further research inclusive of a larger number of speech communities may provide different output. Such studies may yield results in which not only the respective parameters, but additional parameters such as ethnicity, profession, socioeconomic status, and geographical location may be among demographics making statistically significant contributions to language identification in Iran.

Considering the objectives of this research project, this study is neither representative of the entire set of parameters in identifying language varieties nor has it been a practical objective to introduce all the influential parameters. As literature has confirmed abundantly, should the participants' demographic information change, the results may be significantly different. It is also noteworthy that in any given speech community individuals abide by complicated sociocultural norms. Thus, their demographics such as gender and age cannot exist outside of the norms established in their speech community. This is not to say that people have to be in agreement and under a similar influence regarding how one should define or use language as a social entity (see Saville-Troike, 2003). Rather, one needs to consider that, similar to gender, other demographic categories are socially constructed and established.

Overall results of this study indicate that a socially oriented study of linguistic analysis grounded in folk knowledge would provide a more detailed and nuanced perspective of how people in Iran view the language varieties they know, are exposed to or are aware of as part of Iranian national identity. An expansion of understanding may, indeed, facilitate a smoother coexistence of group identities within the nation and promote greater tolerance for variations beyond the established or sanctioned norms. For example, raising folk awareness in this regard may facilitate the adoption of language varieties practiced in speech communities as the medium of education, or at least acknowledged as important companions to the state language in support of all communities within the state.

A folk approach to defining and distinguishing different types of language varieties can raise awareness regarding the values associated with language varieties among speech communities. Folk awareness can further bring attention towards the politically geared categorization of language varieties which tend to classify languages with Farsi. Following government policies, Farsi has been designated as the primary language variety in terms of acceptability and suitability within the government, education systems and social media. However, because of this primacy of one language, people have been compelled to seek social and cultural prestige by speaking only Farsi, often to the detriment of their native language variety which may not only be devalued and downgraded as a language but even negatively pressured to the point of extinction. An inclusive approach to languages in Iran must begin with an understanding of how people actually view their native language and those of others, perspectives for which this study

has attempted to lay the foundation through the development of a useful instrument for relevant data collection.

To address the integration and assimilation policies of the government and instead promote the equal chance of survival for all language varieties, over the years, movements fighting for the rights of minorities and their language varieties have been formed (see Hassanpour, 1994). Such movements aim to expand folk understanding of individual group linguistic rights, preserve language varieties and uphold their status within the nation, not by denying the primacy of one national language but by maintaining diverse legitimate linguistic identities within the state. One way to support a multilinguistic acceptance of such diversity is through research on how folk actually perceive and understand languages which differ from their mother tongue and to analyze how such perceptions arise and are expressed.

Furthermore, recognition of minority languages and their distinctive sociocultural priorities have long been among the goals of minority scholars (see Hassanpour, 1994, 2000; Sheyholislami, 2015, 2019; Skutnabb-Kangas et al., 2012). Seeking folk perspectives of these minority varieties can elaborate on the respective priorities and their influence on how people define and differentiate language varieties. Such folk studies can provide significant details about political pressures and linguicide which have, for years interrupted interdialectal communications, hindered publication in and use of minority languages in educational settings, and banned the formal education of children in their own mother tongue. Though not all influential sociocultural parameters are developed into the language identification questionnaire, qualitative studies can shed light on minority folk perspectives. In addition, investigating folk perspectives about language varieties,

either their own or others', can facilitate discussion with language policy makers regarding language maintenance and potentially come up with approaches to not only preserve and maintain but also prevent the decline and endangerment of minority languages in the society.

Considering the results of this study as well as the available literature on the lack of an official status for language varieties other than the official language of the country (see Section 2.1.4.), investigating linguistic perspectives of minority language speakers about the status of these languages can motivate the minority scholars themselves who have, for decades, observed and critiqued the lack of a recognized status for their languages. Among minority scholars, Sheyholislami, for example, has been raising awareness and teaching in and via Kurdish as a minority language. Such efforts may in turn impact the majority Farsi scholars who continue to define the minority languages as regional and local dialects or varieties of Farsi in their academic publications. By seeking linguistic perspectives, whether folk or scholars', and raising their awareness about the status and vitality of minority languages, languages of Iran should be able to coexist without having to consider one (i.e., Farsi) as the superior language or the others (i.e., minority languages) as threatening the authenticity of Farsi. Should all the languages spoken in Iran exist in harmony, the speakers of minority languages can have language rights without having to assimilate or resist pressures to reject their heritage languages.

Lastly, to continue addressing the issues of language identification as captured in this research project requires a new level of effort and commitment on the part of scholars of Iranian linguistics to consider the realities of language varieties alongside their linguistic features. To move away from extractive projects in their very basic sense where folk are treated only as

linguistic informants, scholars need to learn different ways of thinking about and practicing research, participate in the process of consultation, learn how to be open-minded to folk assumptions and values, and negotiate contradictory expectations and demands. This is where scholars may need to re-evaluate their standard academic and methodological stances. It is worthy of note that juxtaposition of various language varieties including minority languages with the official state language (i.e., Farsi) would also require folk open-mindedness in dealing with linguistic disagreements. Equally important in conducting language-related research is to consider speech communities as equal members while planning the entire research process. The first step in this process is to create methodologies and mechanisms for data collection which focus on the sociocultural aspects of language identification rather than simply formal features.

6.1. Limitations and future research

A key component into the investigation of folk perspectives about languages in Iran was the development of a detailed questionnaire to tease out these views. Considering that this questionnaire is among the first measures to address the issues of language identification in Iran, this tool may lend itself to modification and improvement as it is used in future research. Should future research administer the questionnaire in a different context, researchers may need to re-assess and re-frame the reliability and validity indices. That said, I believe that the questionnaire results can be generalized to many situations and groups, depending on a diverse and numerically sufficient pool of participants from each demographic subgroup.

The most significant challenge in generating the questionnaire was the initial selection of influential parameters associated with people's perspectives. Taking into account all the potential

parameters, even if possible, would result in a complicated questionnaire. As such, recruiting participants willing to spend the time and fill out such a questionnaire would be challenging for researchers (Shultz et al., 2014). Therefore, the components mentioned in the questionnaire should be limited to what would most clearly generate relevant results based on the qualitative data results, panel expertise and a review of the available literature on language identification. Furthermore, adding or deleting items and/or components in this questionnaire may influence construct validity.

Producing quantitative information about constructs that exist under the influence of or in association with numerous influential parameters may require researchers to operationalize the abstract concepts and relationships differently. This is to say that researchers may need to modify the questionnaire to measure linguistic perspectives of people who belong to a given speech community under study. As a solution, future research may benefit from multiple data collection tools including a prior pilot, as did this study.

Even though the findings of this study brought researchers one significant step closer to understanding folk perspectives in language identification, an important avenue of future study could investigate the modification of the questionnaire via the addition or deletion of components and with respect to speech communities under study. Though found to be sensitive in this research project, one such component may investigate whether type and length of exposure to one's own and others' language varieties can influence folk linguistic perspectives.

In the next stages of research, I intend to use this questionnaire as a basis and conduct multiple comparative studies where folk perspectives within and across speech communities from different backgrounds are investigated. One such study can investigate whether folk from the

varieties that exist within different parts of a city (e.g., Tehrani and Jonoub-Shahri Tehrani in the capital city of Tehran) define language variety types differently. Further, I aim to examine perspectives of linguists versus non-linguists regarding these issues. Additionally, I am interested in examining how Iranian linguists' and Western linguists' responses may differ in this matter.

Moreover, among future research objectives are mixed methods studies where a qualitative strand can be conducted as a follow-up phase to using the questionnaire. One such promising line of inquiry is an explanatory study with additional qualitative analysis of Iranian folk definition of *zabān*, *guyesh* and *lahjeh*. The purpose of conducting this study would be two-fold. Firstly, by conducting individual and/or group interviews, I can explore underlying folk reasons for the way they define and differentiate the language variety types in the questionnaire. Secondly, a follow-up qualitative study would provide an opportunity to investigate the implications of all of the participants' Likert-scale choices including 'don't know'.

An additional line of research which seemed interesting from the early stages relates to C2.5, C5.1, and C5.2 of the questionnaire. Though these items turned out to be statistically significant with high factor loadings, I intend to conduct interviews during which the participants can explain why/how a language variety with a writing system may not be *zabān* (C5.1). Further, considering the mixed perspectives about whether *guyesh* has a writing system (C5.2), interviews can elaborate on the underlying reasons for these views. Although a large number of participants from the pilot iterations and administration of the validated questionnaire chose 'don't know' in C2.5, I will conduct interviews to further examine why one would urge the need for a linguistic label to account for language varieties which are "neither *zabān* nor *guyesh*".

As mentioned earlier (see Step four), the participants were given the ‘don’t know’ choice to freely express a neutral response. However, neither during the pilot iterations nor in the administration of the validated questionnaire were these choices investigated further. As Sturgis et al. (2014) assert, the dearth of knowledge regarding the psychological mechanisms that participants use to respond to Likert-scale questions contributes to the difficulty of data interpretation.

Volunteering a ‘don’t know’ response may be a matter of avoiding public admission of ignorance especially in the case of educated participants (Sturgis et al., 2014). Sturgis et al. (2014) further explain that choosing a ‘don’t know’ response could be indicative of motivated participants who genuinely have no knowledge or opinion with regard to the value in question or indicating participants who did not understand the value in question altogether. A ‘don’t know’ option could also point to participants whose opinions are not covered among the available options, and those to whom the value in question seems irrelevant to their life and as such not interesting to them. To these participants, ‘don’t know’ is a cognitively easier alternative. Statistically speaking, a middle or uncommitted response does not guarantee a clearer interpretation of the participants’ motives in choosing that option over others unless follow-up interviews are conducted (Sturgis et al., 2014). Unless the correlation between participants’ demographics and their choice of ‘don’t know’ can be meaningfully established, such interpretation is challenging. It could also be true that those who select a neutral option from among the directional alternative do so only as a face-saving mechanism. In a subsequent paper, I will conduct interviews to examine the reasons underlying

people's decision in selecting the 'don't know' Likert-scale option regarding language identification.

In Step eleven, significant demographic associations were only established with respect to education, formal linguistic knowledge and mother tongue. As part of their demographics, the participants were asked not only about their mother tongue, but also about the speech communities they were born into and speech communities where they lived. In all three phases of this study, there were some participants whose mother tongue, first language and the language variety spoken in the speech community in which they lived were not the same. Therefore, how their perspectives are affected due to the respective parameters requires a separate research paper.

6.2. Final reflections

This dissertation presents some initial steps towards developing this broader understanding, primarily through the development and testing of questionnaires and interviews which elicit folk linguistic views of language, dialect and accent in an attempt to understand and address the issues of language identification. This dissertation is novel in the field of Iranian linguistics, since it is one of the first studies which recognizes language users' perspectives as both accessible and indispensable in language identification. Further, it is novel in developing and validating a language identification questionnaire originated in the context of Iranian speech communities. Since this questionnaire was developed using lay language and based on folk linguistic perspectives, the participants were not challenged by the absence of technical linguistic knowledge. This dissertation further offers new insights about intertwined parameters identified as statistically significant in addressing the issues of language identification. Lastly, the novelty of

this dissertation extends to identifying the extent of variation in linguistic perspectives due to the predictive power of the prominent parameters associated with folk linguistic perspectives.

Appendices

Appendix A

Semi-structured interview questions were developed and organized based on Kvale's (1996) proposed interview type questions (pp. 133-135). Thus, the interviews included introducing, follow-up, probing, specifying, direct, indirect, structuring, interpreting, and throwaway question types. Periods of silence were maintained in between questions in case the interviewees needed to elaborate on an idea.

Demographic questions

- 1) Can you tell me a bit about yourself?
- 2) Where were you born and how old are you?
- 3) Where are your parents from?
- 4) What is your mother tongue?
- 5) Is that your first language?
- 6) Do you know any other language varieties?

In case of those who know/speak more than one language variety:

- 7) Can you speak all these language varieties?
- 8) How about reading and writing? Can you read and write in them?
- 9) How/why did you learn to speak/read/write these language varieties?
- 10) How do you choose which language variety to speak with?
- 11) [For clarity] Is there a priority based on which you switch between the varieties that you speak?

- 12) Can you think of a situation/topic/concept where you switched between two language varieties? [example of situation/topic/concept was provided upon request]
- 13) Which language variety is your favorite TV show or music in?
- 14) Which language variety do you use for social media platforms like Facebook or Instagram?
- 15) Which language variety do you use for your personal matters such as praying, self-talk or thinking out loud?
- 16) Do you have a sense of belonging to any of the varieties that you speak? Do you ever think one or more of them is part of who you are?
- 17) Do you think it is possible to have a sense of belonging to more than one language variety?
[examples provided upon request]
- 18) If yes, why do you think that is? How is it possible?
- 19) What is the language variety in which your parents chose to communicate with you?
- 20) Do you think it was important to you/your parents that you learn and speak your mother tongue?
- 21) Was it important to you/your parents that you learn Farsi too?
- 22) Would it affect your life in any way if you could not speak Farsi?
- 23) Is your mother tongue the language variety that you use when you talk to your friends or family members?
- 24) How about when you speak with your colleagues or manager/boss? Do you communicate using your mother tongue?
- 25) What language variety did you use at school when you were a student?

26) Have you ever had teachers at school who did not speak the language variety that you spoke at home?

27) If yes, what language variety did you use to communicate in class with your teachers?

28) How about exam situations? Were the exams in the language that you spoke at home?

29) Do you think people switch between language varieties in different occasions and under different circumstances?

In case of those who have lived in multiple places:

30) Where have you lived?

31) How long have you lived in each of these places?

32) Since you have lived in multiple places in/outside Iran, which place would you say you belong with?

33) Did moving to a new place affect the choice of language in which you communicated at home or work?

Structuring questions: Now I am going to move to another category of questions which investigates your opinion about what language varieties are. I just want to know your personal point of view, so there is no right or wrong answer.

34) What kind of language variety would you say Farsi or Arabic is?

35) How do you define ... [the variety mentioned in the previous question]?

36) If someone who speaks Farsi and someone who speaks Arabic talk to one another, do you think they would understand each other? Why is that?

- 37) Do you think Isfahani or Shirazi is a ... [the language variety mentioned in answer to the previous question] too? Why?
- 38) If two people, one from Isfahan and one from Shiraz, speak to each other, do you think they understand each other? Why is that?
- 39) How do you think Isfahani or Shirazi is different from your language variety?
- 40) Do you think you would be able to understand someone who speaks either of these varieties? Why?
- 41) How about Bushehri? Do you know what kind of language variety it is?
- 42) Do you think you would be able to understand Bushehri? Why?
- 43) How is Bushehri different from or similar to your language variety?
- 44) What do you think about Kurdish? What kind of language variety is it?
- 45) Do you know what type of language variety Abade'i is? Have you heard of this variety before?
- In case of those participants who differentiated the above-mentioned language varieties:**
- 46) What kind of language variety do you think is used in the education or government? Why?
- 47) In your opinion, is there any difference between a zabān and a guyesh?
- 48) How do you decide what a zabān or a guyesh is?
- 49) Do you think the speakers of two separate zabān would be able to understand each other? [at this point, in case of the interviewees who differentiated Farsi and Arabic as two zabān, I pointed to these two zabān from question #34]
- 50) How about the speakers of two guyesh or two lahjeh? Do you think they would understand one another?

- 51) If someone speaks with you in their own variety and you are only partially able to understand them, what kind of language variety would you say it is?
- 52) Do you think if you know the words of a language variety, you would be able to understand it?
- 53) If you know a lot of words from another language variety, does it mean that this variety is a zabān, guyesh, or lahjeh?
- 54) How about if you know only a few of their words?
- 55) What kind of language variety has a writing system?
- 56) Do you think if a language variety is spoken in a city or town near you, you would be able to understand it?
- 57) What are the chances of understanding a variety if it is spoken somewhere far from you?
- 58) If I tell you that a variety is spoken in a big city like Isfahan or Tehran, would you say it is a zabān, guyesh, or lahjeh? Would you change your answer if I say the variety is spoken in a village instead?

Final questions which seek your understanding of different language varieties:

- 59) If a language variety has a million speakers and more, what kind of language variety would you say it is?
- 60) Do you think guyesh has fewer or more speakers compared with zabān?
- 61) Do you think your views could be different if you had not learned to speak multiple language varieties?

Appendix B

Language identification questionnaire: Languages of Iran

Parisa Sabethemmatbadi (parisasabethemmataba@cmail.carleton.ca)
and Erik Anonby (erik.anonby@carleton.ca)

This questionnaire asks some questions about your views about language varieties. The questionnaire is anonymous. It will not be used to identify you. Do you give permission for your responses to be used in research on language in Iran?

yes no

Preliminary questions

- 1) What is your mother tongue? _____
- 2) What other language varieties (*zabān*, *guyesh*, or *lahjeh*) do you speak?

I don't speak any other language varieties.

I speak the following other language varieties (also mark how well you speak them):

_____ fluently well some a little

3) What is the highest level of education you have completed?

Level of education	In progress	Completed
No formal education	<input type="checkbox"/>	<input type="checkbox"/>
Elementary school	<input type="checkbox"/>	<input type="checkbox"/>
Secondary school	<input type="checkbox"/>	<input type="checkbox"/>
College	<input type="checkbox"/>	<input type="checkbox"/>
Bachelors	<input type="checkbox"/>	<input type="checkbox"/>
Masters	<input type="checkbox"/>	<input type="checkbox"/>
PhD	<input type="checkbox"/>	<input type="checkbox"/>

4) Have you taken any courses in linguistics? _____

5) Gender: _____

6) Year of birth: _____

7) What country were you born in? _____

8) What city or town were you born in? _____

9) Where have you lived? How long did you live in each of the places you resided? At what ages did you live in the respective places?

Country/city	For how long	From what age

For the following 21 statements, please choose the option that best describes your perspectives about language varieties (languages, dialects and accents).

Statements	Agree	Somewhat agree	Somewhat disagree	Disagree	Don't know
People in villages generally speak using dialects.	<input type="checkbox"/>				
People in villages generally speak with accents.	<input type="checkbox"/>				
People in cities generally speak languages.	<input type="checkbox"/>				
A dialect is always part of a larger language because it has fewer speakers.	<input type="checkbox"/>				
If a language variety has only a few thousand speakers, it should generally be considered a dialect.	<input type="checkbox"/>				
If a language variety has at least a million speakers, it should generally be considered a language.	<input type="checkbox"/>				
Dialect is a language variety which, according to the government, has no official status.	<input type="checkbox"/>				
Dialect is a language variety that is not used in the education system.	<input type="checkbox"/>				
Accent is a language variety that is not used in the education system.	<input type="checkbox"/>				
Accent is a language variety which, according to the government, has no official status.	<input type="checkbox"/>				
There should be a linguistic label between 'language' and 'dialect' to account for all the language varieties that are neither a language nor a dialect.	<input type="checkbox"/>				
If two people cannot understand each other when they speak their own language varieties, these varieties should be considered separate languages.	<input type="checkbox"/>				

I cannot understand a language variety if I am not familiar with its words.	<input type="checkbox"/>				
If someone speaks to me in an accent of my mother tongue which is different from my own accent, it is always possible to understand them quite well.	<input type="checkbox"/>				
If someone speaks to me in a dialect of my mother tongue which is different from my own dialect, it is always possible to understand them quite well.	<input type="checkbox"/>				
If two language varieties share less than 50% of their words, they are generally separate languages.	<input type="checkbox"/>				
If two language varieties share more than 80% of their words, they are generally accents of the same language.	<input type="checkbox"/>				
If two language varieties share more than 80% of their words, they are generally dialects of the same language.	<input type="checkbox"/>				
If a language variety has a standardized writing system, it may not be a language.	<input type="checkbox"/>				
Dialects do not have standardized writing systems.	<input type="checkbox"/>				
To be considered a language, a language variety needs to have a standardized writing system.	<input type="checkbox"/>				

Please share your feedback on this questionnaire with us.

پرسشنامه شناسایی زبان: زبان های ایران

پریسا ثابت همت‌آبادی (parisasabethemmataba@gmail.com) (parisasabethemmataba@carleton.ca)

و اریک آنونبی (erik.anonby@carleton.ca)

نسخه: ۱۳۹۹/۱۲/۲۶

این پرسشنامه جویای دیدگاه‌های شما درباره‌ی گونه‌های زبانی است. پرسشنامه به صورت ناشناس پر خواهد شد و برای شناسایی شما استفاده نخواهد شد. آیا اجازه می‌دهید که پاسخ‌های شما برای تحقیق در مورد گونه‌های زبانی مورد استفاده قرار

گیرد؟ بله خیر

پرسش های مقدماتی

- زبان مادری (زبان اول) شما چیست؟ _____
- به چه گونه‌های زبانی (زبان، گویش و یا لهجه) دیگری صحبت می‌کنید؟
 من به هیچ گونه‌ی زبانی دیگری صحبت نمی‌کنم.
 من به گونه‌های زبانی زیر صحبت می‌کنم (همچنین تعیین کنید که به هر کدام تا چه حد تسلط دارید):
 _____ روان خوب تا اندازه ای کمی
- _____ روان خوب تا اندازه ای کمی

● بالاترین سطح تحصیلات رسمی شما چیست؟

سطح تحصیلات	در حال انجام	تکمیل شده
بدون آموزش رسمی	<input type="checkbox"/>	<input type="checkbox"/>
ابتدایی	<input type="checkbox"/>	<input type="checkbox"/>
متوسطه	<input type="checkbox"/>	<input type="checkbox"/>
فوق دیپلم	<input type="checkbox"/>	<input type="checkbox"/>
لیسانس	<input type="checkbox"/>	<input type="checkbox"/>
فوق لیسانس	<input type="checkbox"/>	<input type="checkbox"/>
دکتری	<input type="checkbox"/>	<input type="checkbox"/>

● آیا تا به حال درسی در زبانشناسی گذرانده‌اید؟ بله خیر

● جنسیت: _____

● سال تولد: _____

● در کدام کشور متولد شدید؟ _____

● در کدام شهر متولد شدید؟ _____

• تا به حال در کدام کشورها/ شهرها زندگی کرده‌اید؟ چه مدت در هر کشور/ شهری که در آن اقامت داشته‌اید زندگی کردید؟

در چه سنی در مکان های ذکر شده زندگی کرده‌اید ؟

کشور/ شهر	مدت اقامت	از چه سنی

در مورد هر یک از جملات زیر، لطفاً گزینه ای را انتخاب کنید که دیدگاه شما را در مورد گونه‌های زبانی به بهترین وجه توصیف کند.

موافقم	تا حدودی موافقم	تا حدودی مخالفم	مخالفم	نمیدانم	
<input type="checkbox"/>	مردم در روستاها معمولاً به گویش صحبت می کنند.				
<input type="checkbox"/>	مردم در روستاها معمولاً به لهجه صحبت می کنند.				
<input type="checkbox"/>	مردم در شهرها معمولاً به زبان صحبت می کنند.				
<input type="checkbox"/>	یک گویش همیشه بخشی از یک زبان بزرگتر است زیرا تعداد کمتری سخنور دارد.				
<input type="checkbox"/>	اگر یک گونه زبانی فقط چند هزار سخنور داشته باشد، به طور کلی باید آن گونه را یک گویش در نظر گرفت.				
<input type="checkbox"/>	اگر یک گونه زبانی حداقل یک میلیون سخنور داشته باشد، به طور کلی باید آن گونه را یک زبان در نظر گرفت.				
<input type="checkbox"/>	گویش گونه‌ای از زبان است که از نظر دولت به رسمیت شناخته نشده است.				
<input type="checkbox"/>	گویش گونه‌ای از زبان است که در سیستم آموزش و پرورش مورد استفاده قرار نمی‌گیرد.				
<input type="checkbox"/>	لهجه گونه ای از زبان است که در سیستم آموزش و پرورش مورد استفاده قرار نمی‌گیرد.				
<input type="checkbox"/>	لهجه گونه ای از زبان است که از نظر دولت به رسمیت شناخته نشده است.				

<input type="checkbox"/>	باید یک اصطلاح زبانی بین زبان و گویش وجود داشته باشد تا با استفاده از آن به گونه‌های زبانی که نه زبان و نه گویش محسوب می‌شوند اشاره کرد.				
<input type="checkbox"/>	اگر دو نفر هنگام صحبت کردن به گونه‌های زبانی خود، صحبت یکدیگر را نفهمند، این گونه‌های زبانی باید زبان‌های جداگانه تلقی شوند.				
<input type="checkbox"/>	اگر با کلمات یک گونه‌ی زبانی آشنا نباشم، نمی‌توانم آن گونه‌ی زبانی را بفهمم.				
<input type="checkbox"/>	اگر کسی به لهجه‌ای از زبان مادری من که متفاوت از لهجه من است صحبت کند، همیشه می‌توانم صحبت آنها را کاملاً بفهمم.				
<input type="checkbox"/>	اگر کسی به گویشی از زبان مادری من که متفاوت از گویش من است صحبت کند، همیشه می‌توانم صحبت آنها را کاملاً بفهمم.				
<input type="checkbox"/>	اگر دو گونه‌ی زبانی مختلف، کمتر از ۵۰٪ کلمات مشترک داشته باشند، معمولاً دو زبان جداگانه تلقی می‌شوند.				
<input type="checkbox"/>	اگر دو گونه‌ی زبانی مختلف، بیشتر از ۸۰٪ کلمات مشترک داشته باشند، معمولاً دو لهجه از یک زبان تلقی می‌شوند.				
<input type="checkbox"/>	اگر دو گونه‌ی زبانی مختلف، بیشتر از ۸۰٪ کلمات مشترک داشته باشند، معمولاً دو گویش از یک زبان تلقی می‌شوند.				
<input type="checkbox"/>	اگر یک گونه‌ی زبانی دارای یک سیستم نوشتاری استاندارد باشد، ممکن است یک زبان نباشد.				
<input type="checkbox"/>	گویش‌ها سیستم نوشتاری استاندارد ندارند.				
<input type="checkbox"/>	برای اینکه یک زبان تلقی شود، یک گونه‌ی زبانی باید سیستم نوشتاری استاندارد داشته باشد.				

لطفاً نظرات خود در مورد این پرسشنامه را با ما به اشتراک بگذارید.

Appendix C

In this Appendix details regarding the speech community members' demographics during the pilot iterations, EFA, CFA, and administration of the validated questionnaire are presented.

Second pilot iteration

Table 79. Farsi speech community members in the second pilot iteration

		Formal linguistic knowledge		Education				Age				Gender	
		Yes	No	High school	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Male	Female
Farsi speech community members	Tehran	2	5	0	3	4	0	2	2	0	3	2	5
		28.6%	71.4%	0.0%	42.9%	57.1%	0.0%	28.6%	28.6%	0.0%	42.9%	28.6%	71.4%
	Kermanshah	0	3	0	0	3	0	0	0	0	3	1	2
		0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	33.3%	66.7%
	Bushehr	0	7	4	0	1	2	4	0	3	0	2	5
		0.0%	100.0%	57.1%	0.0%	14.3%	28.6%	57.1%	0.0%	42.9%	0.0%	28.6%	71.4%
	Shiraz	0	4	0	0	2	2	0	2	2	0	1	3
	0.0%	100.0%	0.0%	0.0%	50.0%	50.0%	0.0%	50.0%	50.0%	0.0%	25.0%	75.0%	
Zahedan	0	4	4	0	0	0	4	0	0	0	2	2	
	0.0%	100.0%	100.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	50.0%	50.0%	
Total	2	23	8	3	10	4	10	4	5	6	8	17	
	8.0%	92.0%	32.0%	12.0%	40.0%	16.0%	40.0%	16.0%	20.0%	24.0%	32.0%	68.0%	

Table 80. Kurdish speech community members in the second pilot iteration

		Formal linguistic knowledge		Education				Age				Gender	
		Yes	No	High school	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Male	Female
Kurdish speech community members	Ilam	0	6	0	0	4	2	0	6	0	0	3	3
		0.0%	100.0%	0.0%	0.0%	66.7%	33.3%	0.0%	100.0%	0.0%	0.0%	50.0%	50.0%
	Mahabad	0	4	1	3	0	0	1	0	0	3	2	2
	0.0%	100.0%	25.0%	75.0%	0.0%	0.0%	25.0%	0.0%	0.0%	75.0%	50.0%	50.0%	
Total	0	10	1	3	4	2	1	6	0	3	5	5	
	0.0%	100.0%	10.0%	30.0%	40.0%	20.0%	10.0%	60.0%	0.0%	30.0%	50.0%	50.0%	

Table 81. Turkic speech community members in the second pilot iteration

		Formal linguistic knowledge		Education				Age				Gender	
		Yes	No	High school	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Male	Female
Turkic speech community members	Tabriz	0	6	0	0	6	0	0	6	0	0	4	2
		0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	66.7%	33.3%
	Zanjan	4	2	0	0	2	4	0	6	0	0	1	5
	66.7%	33.3%	0.0%	0.0%	33.3%	66.7%	0.0%	100.0%	0.0%	0.0%	16.7%	83.3%	
Total	4	8	0	0	8	4	0	12	0	0	5	7	
	33.3%	66.7%	0.0%	0.0%	66.7%	33.3%	0.0%	100.0%	0.0%	0.0%	41.7%	58.3%	

Table 82. Mazandarani speech community members in the second pilot iteration

		Formal linguistic knowledge		Education				Age				Gender	
		Yes	No	High school	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Male	Female
Mazandarani speech community members	Sari	3	2	0	0	5	0	0	5	0	0	2	3
		60.0%	40.0%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	40.0%	60.0%
	Amol	0	4	0	0	4	0	0	4	0	0	1	3
		0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	25.0%	75.0%
	Total	3	6	0	0	9	0	0	9	0	0	3	6
33.3%		66.7%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	33.3%	66.7%	

Third pilot iteration

Table 83. Farsi speech community members in the third pilot iteration

		Formal linguistic knowledge		Education			Age			Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	Male	Female
Farsi speech community members	Tehran	7 23.3%	23 76.7%	2 6.7%	22 73.3%	6 20.0%	7 23.3%	15 50.0%	8 26.7%	10 33.3%	20 66.7%
	Yazd	0 0.0%	3 100.0%	0 0.0%	3 100.0%	0 0.0%	0 0.0%	3 100.0%	0 0.0%	2 66.7%	1 33.3%
	Mashhad	1 20.0%	4 80.0%	0 0.0%	5 100.0%	0 0.0%	1 20.0%	4 80.0%	0 0.0%	1 20.0%	4 80.0%
	Semnan	2 100.0%	0 0.0%	0 0.0%	2 100.0%	0 0.0%	0 0.0%	2 100.0%	0 0.0%	2 100.0%	0 0.0%
	Kerman	0 0.0%	2 100.0%	0 0.0%	2 100.0%	0 0.0%	0 0.0%	2 100.0%	0 0.0%	2 100.0%	0 0.0%
	Total	10 23.8%	32 76.2%	2 4.8%	34 81.0%	6 14.3%	8 19.0%	26 61.9%	8 19.0%	17 40.5%	25 59.5%

Table 84. Kurdish speech community members in the third pilot iteration

		Formal linguistic knowledge		Education			Age			Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	Male	Female
Kurdish speech community members	Abdanan	0 0.0%	1 100.0%	0 0.0%	1 100.0%	0 0.0%	0 0.0%	1 100.0%	0 0.0%	1 100.0%	0 0.0%
	Urmia	0 0.0%	4 100.0%	0 0.0%	4 100.0%	0 0.0%	2 50.0%	2 50.0%	0 0.0%	0 0.0%	4 100.0%
	Ilam	1 20.0%	4 80.0%	0 0.0%	5 100.0%	0 0.0%	0 0.0%	5 100.0%	0 0.0%	5 100.0%	0 0.0%
	Total	1 10.0%	9 90.0%	0 0.0%	10 100.0%	0 0.0%	2 20.0%	8 80.0%	0 0.0%	6 60.0%	4 40.0%

Table 85. Turkic speech community members in the third pilot iteration

		Formal linguistic knowledge		Education			Age			Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	Male	Female
Turkic speech community members	Tabriz	0 0.0%	7 100.0%	0 0.0%	7 100.0%	0 0.0%	2 28.6%	5 71.4%	0 0.0%	0 0.0%	7 100.0%
	Arak	2 100.0%	0 0.0%	0 0.0%	2 100.0%	0 0.0%	0 0.0%	2 100.0%	0 0.0%	2 100.0%	0 0.0%
	Salmas	1 100.0%	0 0.0%	0 0.0%	1 100.0%	0 0.0%	0 0.0%	1 100.0%	0 0.0%	1 100.0%	0 0.0%
	Meshginshahr	0 0.0%	2 100.0%	0 0.0%	2 100.0%	0 0.0%	0 0.0%	2 100.0%	0 0.0%	1 50.0%	1 50.0%
	Total	3 25.0%	9 75.0%	0 0.0%	12 100.0%	0 0.0%	2 16.7%	10 83.3%	0 0.0%	4 33.3%	8 66.7%

Table 86. Gilaki speech community members in the third pilot iteration

		Formal linguistic knowledge		Education			Age			Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	Male	Female
Gilaki speech community members	Rasht	0 0.0%	4 100.0%	0 0.0%	4 100.0%	0 0.0%	2 50.0%	2 50.0%	0 0.0%	0 0.0%	4 100.0%
	Total	0 0.0%	4 100.0%	0 0.0%	4 100.0%	0 0.0%	2 50.0%	2 50.0%	0 0.0%	0 0.0%	4 100.0%

Table 87. Lori speech community members in the third pilot iteration

		Formal linguistic knowledge		Education			Age			Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	Male	Female
Lori speech community members	Borujerd	1	1	0	2	0	0	2	0	0	2
		50.0%	50.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%
	Total	1	1	0	2	0	0	2	0	0	2
		50.0%	50.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%

Exploratory Factor Analysis (EFA)

Table 88. Farsi speech community members during the Exploratory Factor Analysis (EFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Farsi speech community members	Tehran	2	25	16	7	4	2	6	4	8	7	16	11
		7.4%	92.6%	59.3%	25.0%	14.8%	7.4%	22.2%	14.8%	29.6%	25.9%	59.3%	40.7%
	Shiraz	1	11	2	7	3	1	7	3	1	0	3	9
		8.3%	91.7%	16.7%	58.3%	25.0%	8.3%	58.3%	25.0%	8.3%	0.0%	25.0%	75.0%
	Isfahan	1	12	2	10	1	0	1	5	4	3	9	4
		7.7%	92.3%	15.4%	76.9%	7.7%	0.0%	7.7%	38.5%	30.8%	23.1%	69.2%	30.8%
	Mashhad	0	10	2	6	2	2	1	2	3	2	1	9
		0.0%	100.0%	20.0%	60.0%	20.0%	20.0%	10.0%	20.0%	30.0%	20.0%	10.0%	90.0%
	Kermanshah	1	7	2	5	1	0	3	4	1	0	1	7
		12.5%	87.5%	25.0%	62.5%	12.5%	0.0%	37.5%	50.0%	12.5%	0.0%	12.5%	87.5%
	Kerman	0	5	1	4	0	0	0	1	4	0	3	2
		0.0%	100.0%	20.0%	80.0%	0.0%	0.0%	0.0%	20.0%	80.0%	0.0%	60.0%	40.0%
	Bushehr	0	4	0	4	0	0	4	0	0	0	0	4
		0.0%	100.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%
Yazd	0	2	2	0	0	0	0	0	1	1	2	0	
	0.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%	50.0%	100.0%	0.0%	
Semnan	0	2	2	0	0	0	0	0	2	0	2	0	
	0.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	100.0%	0.0%	
Total	5	78	29	43	11	5	22	19	24	13	37	46	
	6.0%	94.0%	34.9%	51.8%	13.3%	6.0%	26.5%	22.9%	28.9%	15.7%	44.6%	55.4%	

Table 89. Kurdish speech community members during the Exploratory Factor Analysis (EFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Kurdish speech community members	Sanandaj	3	2	0	3	2	0	2	2	1	0	1	4
		60.0%	40.0%	0.0%	60.0%	40.0%	0.0%	40.0%	40.0%	20.0%	0.0%	20.0%	80.0%
	Ilam	0	8	2	5	1	1	1	5	1	0	2	6
		0.0%	100.0%	25.0%	62.5%	12.5%	12.5%	12.5%	62.5%	12.5%	0.0%	25.0%	75.0%
Total	3	10	2	8	3	1	3	7	2	0	3	10	
	23.1%	76.9%	15.4%	61.5%	23.1%	7.7%	23.1%	53.8%	15.4%	0.0%	23.1%	76.9%	

Table 90. Turkic speech community members during the Exploratory Factor Analysis (EFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Turkic speech community members	Tabriz	5	33	10	22	6	2	12	8	10	6	10	28
		13.2%	86.8%	26.3%	57.9%	15.8%	5.3%	31.6%	21.1%	26.3%	15.8%	26.3%	73.7%
	Ardabil	1	9	6	2	2	0	2	2	3	3	7	3
		10.0%	90.0%	60.0%	20.0%	20.0%	0.0%	20.0%	20.0%	30.0%	30.0%	70.0%	30.0%
	Urmia	2	10	3	7	2	0	5	6	0	1	4	8
		16.7%	83.3%	25.0%	58.3%	16.7%	0.0%	41.7%	50.0%	0.0%	8.3%	33.3%	66.7%
	Tarom	3	7	5	3	2	0	1	1	3	5	6	4
		30.0%	70.0%	50.0%	30.0%	20.0%	0.0%	10.0%	10.0%	30.0%	50.0%	60.0%	40.0%
Total	11	59	24	34	12	2	20	17	16	15	27	43	
	15.7%	84.3%	34.3%	48.6%	17.1%	2.9%	28.6%	24.3%	22.9%	21.4%	38.6%	61.4%	

Table 91. Mazandarani speech community members during the Exploratory Factor Analysis (EFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Mazandarani speech community members	Amol	1	6	2	3	2	2	1	2	2	0	2	5
		14.3%	85.7%	28.6%	42.9%	28.6%	28.6%	14.3%	28.6%	28.6%	0.0%	28.6%	71.4%
	Babol	0	2	0	1	1	0	1	1	0	0	1	1
		0.0%	100.0%	0.0%	50.0%	50.0%	0.0%	50.0%	50.0%	0.0%	0.0%	50.0%	50.0%
	Total	1	8	2	4	3	2	2	3	2	0	3	6
		11.1%	88.9%	22.2%	44.4%	33.3%	22.2%	22.2%	33.3%	22.2%	0.0%	33.3%	66.7%

Table 92. Balochi speech community members during the Exploratory Factor Analysis (EFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Balochi speech community members	Iranshahr	0	5	1	3	1	1	1	2	1	0	3	2
		0.0%	100.0%	20.0%	60.0%	20.0%	20.0%	20.0%	40.0%	20.0%	0.0%	60.0%	40.0%
	Chabahar	0	3	0	3	0	2	1	0	0	0	0	3
		0.0%	100.0%	0.0%	100.0%	0.0%	66.7%	33.3%	0.0%	0.0%	0.0%	0.0%	100.0%
	Total	0	8	1	6	1	3	2	2	1	0	3	5
		0.0%	100.0%	12.5%	75.0%	12.5%	37.5%	25.0%	25.0%	12.5%	0.0%	37.5%	62.5%

Table 93. Lori speech community members during the Exploratory Factor Analysis (EFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Lori speech community members	Yasuj	0	7	3	4	0	0	4	0	2	1	1	6
		0.0%	100.0%	42.9%	57.1%	0.0%	0.0%	57.1%	0.0%	28.6%	14.3%	14.3%	85.7%
	Khorramabad	0	2	1	1	0	0	0	0	1	1	0	2
		0.0%	100.0%	50.0%	50.0%	0.0%	0.0%	0.0%	50.0%	50.0%	0.0%	0.0%	100.0%
	Total	0	9	4	5	0	0	4	0	3	2	1	8
		0.0%	100.0%	44.4%	55.6%	0.0%	0.0%	44.4%	0.0%	33.3%	22.2%	11.1%	88.9%

Table 94. Gilaki speech community members during the Exploratory Factor Analysis (EFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Gilaki speech community members	Rasht	0	5	2	2	1	1	1	1	1	1	0	5
		0.0%	100.0%	40.0%	40.0%	20.0%	20.0%	20.0%	20.0%	20.0%	20.0%	0.0%	100.0%
	Lahijan	1	3	0	2	2	0	2	1	1	0	1	3
		25.0%	75.0%	0.0%	50.0%	50.0%	0.0%	50.0%	25.0%	25.0%	0.0%	25.0%	75.0%
	Fuman	0	2	0	2	0	0	2	0	0	0	0	2
		0.0%	100.0%	0.0%	100.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%
	Total	1	10	2	6	3	1	5	2	2	1	1	10
		9.1%	90.9%	18.2%	54.5%	27.3%	9.1%	45.5%	18.2%	18.2%	9.1%	9.1%	90.9%

Confirmatory Factor Analysis (CFA)

Table 95. Farsi speech community members during the Confirmatory Factor Analysis (CFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Farsi speech community members	Tehran	6	57	22	31	10	3	16	15	19	10	34	29
		9.5%	90.5%	34.9%	49%	15.9%	4.8%	25.4%	23.8%	30.2%	15.9%	54%	46.0%
	Karaj	1	9	6	2	2	1	3	1	3	2	5	5
		10.0%	90.0%	60.0%	20%	20.0%	10.0%	30.0%	10.0%	30.0%	20.0%	50%	50.0%
	Qom	0	7	4	2	1	1	3	0	1	2	2	5
		0.0%	100.0%	57.1%	29%	14.3%	14.3%	42.9%	0.0%	14.3%	28.6%	29%	71.4%
	Ahvaz	1	4	2	2	1	0	1	2	2	0	0	5
		20.0%	80.0%	40.0%	40%	20.0%	0.0%	20.0%	40.0%	40.0%	0.0%	0.0%	100.0%
	Bushehr	0	10	1	9	0	0	7	2	1	0	1	9
		0.0%	100.0%	10.0%	90%	0.0%	0.0%	70.0%	20.0%	10.0%	0.0%	10%	90.0%
	Abadeh	0	5	2	3	0	0	1	0	2	2	3	2
		0.0%	100.0%	40.0%	60%	0.0%	0.0%	20.0%	0.0%	40.0%	40.0%	60%	40.0%
	Shiraz	0	7	2	5	0	0	1	2	4	0	4	3
		0.0%	100.0%	28.6%	71%	0.0%	0.0%	14.3%	28.6%	57.1%	0.0%	57%	42.9%
	Isfahan	1	10	4	4	3	1	3	2	5	0	4	7
		9.1%	90.9%	36.4%	36%	27.3%	9.1%	27.3%	18.2%	45.5%	0.0%	36%	63.6%
	Zahedan	1	1	1	0	1	0	0	0	2	0	2	0
		50.0%	50.0%	50.0%	0.0%	50.0%	0.0%	0.0%	0.0%	100.0%	0.0%	100%	0.0%
	Borazjan	0	2	0	2	0	0	2	0	0	0	2	0
		0.0%	100.0%	0.0%	100%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	100%	0.0%
Kerman	3	0	0	3	0	0	0	2	1	0	0	3	
	100.0%	0.0%	0.0%	100%	0.0%	0.0%	0.0%	66.7%	33.3%	0.0%	0.0%	100.0%	
Kermanshah	1	4	0	4	1	0	4	1	0	0	2	3	
	20.0%	80.0%	0.0%	80%	20.0%	0.0%	80.0%	20.0%	0.0%	0.0%	40%	60.0%	
Semnan	3	2	0	2	3	1	3	0	1	0	3	2	
	60.0%	40.0%	0.0%	40%	60.0%	20.0%	60.0%	0.0%	20.0%	0.0%	60%	40.0%	
Yazd	0	4	3	1	0	0	1	0	0	3	3	1	
	0.0%	100.0%	75.0%	25%	0.0%	0.0%	25.0%	0.0%	0.0%	75.0%	75%	25.0%	
Total	17	122	47	70	22	7	45	27	41	19	65	74	
	12.2%	87.8%	33.8%	50%	15.8%	5.0%	32.4%	19.4%	29.5%	13.7%	47%	53.2%	

Table 96. Kurdish speech community members during the Confirmatory Factor Analysis (CFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Kurdish speech community members	Ilam	3	6	0	7	2	0	2	6	1	0	1	8
		33.3%	66.7%	0.0%	77.8%	22.2%	0.0%	22.2%	66.7%	11.1%	0.0%	11.1%	88.9%
	Sanandaj	5	8	5	4	4	1	3	4	3	2	5	8
		38.5%	61.5%	38.5%	30.8%	30.8%	7.7%	23.1%	30.8%	23.1%	15.4%	38.5%	61.5%
	Baneh	1	2	1	1	1	0	2	1	0	0	1	2
		33.3%	66.7%	33.3%	33.3%	33.3%	0.0%	66.7%	33.3%	0.0%	0.0%	33.3%	66.7%
	Paveh	1	3	1	1	2	0	0	2	2	0	2	2
		25.0%	75.0%	25.0%	25.0%	50.0%	0.0%	0.0%	50.0%	50.0%	0.0%	50.0%	50.0%
	Mahabad	3	18	10	11	0	2	8	4	5	2	8	13
		14.3%	85.7%	47.6%	52.4%	0.0%	9.5%	38.1%	19.0%	23.8%	9.5%	38.1%	61.9%
	Abdanan	1	3	2	1	1	0	2	0	2	0	3	1
		25.0%	75.0%	50.0%	25.0%	25.0%	0.0%	50.0%	0.0%	50.0%	0.0%	75.0%	25.0%
	Total	14	40	19	25	10	3	17	17	13	4	20	34
		25.9%	74.1%	35.2%	46.3%	18.5%	5.6%	31.5%	31.5%	24.1%	7.4%	37.0%	63.0%

Table 97. Turkic speech community members during the Confirmatory Factor Analysis (CFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Turkic speech community members	Tabriz	13	67	26	41	13	2	24	21	18	15	30	50
		16.3%	83.8%	32.5%	51.2%	16.3%	2.5%	30.0%	26.3%	22.5%	18.8%	37.5%	62.5%
	Urmia	7	8	2	7	6	1	4	3	7	0	7	8
		46.7%	53.3%	13.3%	46.7%	40.0%	6.7%	26.7%	20.0%	46.7%	0.0%	46.7%	53.3%
	Ardabil	8	2	0	5	5	0	3	4	3	0	6	4
		80.0%	20.0%	0.0%	50.0%	50.0%	0.0%	30.0%	40.0%	30.0%	0.0%	60.0%	40.0%
	Zanjan	2	11	6	4	3	1	4	4	3	1	5	8
15.4%		84.6%	46.2%	30.8%	23.1%	7.7%	30.8%	30.8%	23.1%	7.7%	38.5%	61.5%	
Total	30	88	34	57	27	4	35	32	31	16	48	70	
	25.4%	74.6%	28.8%	48.3%	22.9%	3.4%	29.7%	27.1%	26.3%	13.6%	40.7%	59.3%	

Table 98. Mazandarani speech community members during the Confirmatory Factor Analysis (CFA)

		Formal linguistic knowledge		Education			Age					Gender	
		Yes	No	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Mazandarani speech community members	Nowshahr	1	13	5	6	3	1	5	3	4	1	3	11
		7.1%	92.9%	35.7%	42.9%	21.4%	7.1%	35.7%	21.4%	28.6%	7.1%	21.4%	78.6%
	Amol	0	10	3	5	2	1	2	3	2	2	3	7
		0.0%	100.0%	30.0%	50.0%	20.0%	10.0%	20.0%	30.0%	20.0%	20.0%	30.0%	70.0%
	Babol	1	10	1	8	2	4	4	1	2	0	1	10
		9.1%	90.9%	9.1%	72.7%	18.2%	36.4%	36.4%	9.1%	18.2%	0.0%	9.1%	90.9%
	Sari	0	15	2	8	5	0	6	9	0	0	5	10
		0.0%	100.0%	13.3%	53.3%	33.3%	0.0%	40.0%	60.0%	0.0%	0.0%	33.3%	66.7%
	Behshahr	5	14	7	6	6	1	6	4	4	4	6	13
		26.3%	73.7%	36.8%	31.6%	31.6%	5.3%	31.6%	21.1%	21.1%	21.1%	31.6%	68.4%
	Babolsar	2	7	2	6	1	2	2	3	2	0	2	7
		22.2%	77.8%	22.2%	66.7%	11.1%	22.2%	22.2%	33.3%	22.2%	0.0%	22.2%	77.8%
	Total	9	69	20	39	19	9	25	23	14	7	20	58
		11.5%	88.5%	25.6%	50.0%	24.4%	11.5%	32.1%	29.5%	17.9%	9.0%	25.6%	74.4%

Speech community members in the administration of the language identification questionnaire

Table 99. Farsi speech community members in the administration of the language identification questionnaire

		Formal linguistic knowledge		Education			Age					Gender	
		No	Yes	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Farsi speech community members	Tehran	6	4	6	2	2	0	2	5	3	0	6	4
		60.0%	40.0%	60.0%	20.0%	20.0%	0.0%	20.0%	50.0%	30.0%	0.0%	60.0%	40.0%
	Mashhad	2	2	1	2	1	0	3	1	0	0	4	0
		50.0%	50.0%	25.0%	50.0%	25.0%	0.0%	75.0%	25.0%	0.0%	0.0%	100.0%	0.0%
	Isfahan	2	1	0	2	1	0	2	1	0	0	3	0
		66.7%	33.3%	0.0%	66.7%	33.3%	0.0%	66.7%	33.3%	0.0%	0.0%	100.0%	0.0%
	Shiraz	2	1	2	0	1	1	1	0	1	0	3	0
		66.7%	33.3%	66.7%	0.0%	33.3%	33.3%	33.3%	0.0%	33.3%	0.0%	100.0%	0.0%
	Kermanshah	4	0	0	3	1	0	3	0	0	1	0	4
		100.0%	0.0%	0.0%	75.0%	25.0%	0.0%	75.0%	0.0%	0.0%	25.0%	0.0%	100.0%
	Ghom	2	0	1	0	1	0	0	1	1	0	2	0
		100.0%	0.0%	50.0%	0.0%	50.0%	0.0%	0.0%	50.0%	50.0%	0.0%	100.0%	0.0%
	Ahvaz	2	1	1	1	1	0	2	0	1	0	0	3
		66.7%	33.3%	33.3%	33.3%	33.3%	0.0%	66.7%	0.0%	33.3%	0.0%	0.0%	100.0%
	Abadeh	2	0	0	0	2	0	1	1	0	0	0	2
		100.0%	0.0%	0.0%	0.0%	100.0%	0.0%	50.0%	50.0%	0.0%	0.0%	0.0%	100.0%
	Bushehr	4	0	0	3	1	1	1	1	1	0	0	4
		100.0%	0.0%	0.0%	75.0%	25.0%	25.0%	25.0%	25.0%	25.0%	0.0%	0.0%	100.0%
	Kerman	3	0	0	2	1	1	1	0	0	1	0	3
		100.0%	0.0%	0.0%	66.7%	33.3%	33.3%	33.3%	0.0%	0.0%	33.3%	0.0%	100.0%
Total	29	9	11	15	12	3	16	10	7	2	18	20	
	76.3%	23.7%	28.9%	39.5%	31.6%	7.9%	42.1%	26.3%	18.4%	5.3%	47.4%	52.6%	

Table 100. Kurdish speech community members in the administration of the language identification questionnaire

		Formal linguistic knowledge		Education			Age					Gender	
		No	Yes	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Kurdish speech community members	Ilam	10	5	8	2	5	0	6	7	2	0	11	4
		66.7%	33.3%	53.3%	13.3%	33.3%	0.0%	40.0%	46.7%	13.3%	0.0%	73.3%	26.7%
	Kermanshah	11	2	6	5	2	2	3	8	0	0	0	13
		84.6%	15.4%	46.2%	38.5%	15.4%	15.4%	23.1%	61.5%	0.0%	0.0%	0.0%	100.0%
	Sanandaj	15	0	9	2	4	0	3	8	4	0	1	14
		100.0%	0.0%	60.0%	13.3%	26.7%	0.0%	20.0%	53.3%	26.7%	0.0%	6.7%	93.3%
	Mahabad	3	3	2	3	1	2	1	2	1	0	2	4
		50.0%	50.0%	33.3%	50.0%	16.7%	33.3%	16.7%	33.3%	16.7%	0.0%	33.3%	66.7%
	Urmia	17	0	9	4	4	0	4	6	7	0	9	8
		100.0%	0.0%	52.9%	23.5%	23.5%	0.0%	23.5%	35.3%	41.2%	0.0%	52.9%	47.1%
	Sardasht	1	3	1	1	2	0	1	3	0	0	1	3
		25.0%	75.0%	25.0%	25.0%	50.0%	0.0%	25.0%	75.0%	0.0%	0.0%	25.0%	75.0%
	Total	57	13	35	17	18	4	18	34	14	0	24	46
		81.4%	18.6%	50.0%	24.3%	25.7%	5.7%	25.7%	48.6%	20.0%	0.0%	34.3%	65.7%

Table 101. Turkic speech community members in the administration of the language identification questionnaire

		Formal linguistic knowledge		Education			Age					Gender	
		No	Yes	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Turkic speech community members	Tabriz	22	15	8	14	15	11	15	7	3	1	20	17
		59.5%	40.5%	21.6%	37.8%	40.5%	29.7%	40.5%	18.9%	8.1%	2.7%	54.1%	45.9%
	Urmia	11	0	3	7	1	3	6	2	0	0	0	11
		100.0%	0.0%	27.3%	63.6%	9.1%	27.3%	54.5%	18.2%	0.0%	0.0%	0.0%	100.0%
	Zanjan	9	0	2	5	2	1	6	2	0	0	0	9
		100.0%	0.0%	22.2%	55.6%	22.2%	11.1%	66.7%	22.2%	0.0%	0.0%	0.0%	100.0%
	Ghazvin	6	1	4	3	0	0	4	2	1	0	6	1
		85.7%	14.3%	57.1%	42.9%	0.0%	0.0%	57.1%	28.6%	14.3%	0.0%	85.7%	14.3%
	Ardabil	8	0	3	3	2	0	0	2	4	2	0	8
		100.0%	0.0%	37.5%	37.5%	25.0%	0.0%	0.0%	25.0%	50.0%	25.0%	0.0%	100.0%
	Total	56	16	20	32	20	15	31	15	8	3	26	46
		77.8%	22.2%	27.8%	44.4%	27.8%	20.8%	43.1%	20.8%	11.1%	4.2%	36.1%	63.9%

Table 102. Mazandarani speech community members in the administration of the language identification questionnaire

		Formal linguistic knowledge		Education			Age					Gender	
		No	Yes	College	BA	MA/PhD	Under 20	20-30	30-40	40-50	Over 50	Male	Female
Mazandarani speech community members	Amol	4	0	0	4	0	2	2	0	0	0	0	4
		100.0%	0.0%	0.0%	100.0%	0.0%	50.0%	50.0%	0.0%	0.0%	0.0%	0.0%	100.0%
	Sari	3	0	0	3	0	1	1	1	0	0	0	3
		100.0%	0.0%	0.0%	100.0%	0.0%	33.3%	33.3%	33.3%	0.0%	0.0%	0.0%	100.0%
	Babol	3	0	0	3	0	1	1	0	0	1	0	3
		100.0%	0.0%	0.0%	100.0%	0.0%	33.3%	33.3%	0.0%	0.0%	33.3%	0.0%	100.0%
	Tonekabon	2	0	0	1	1	0	1	1	0	0	1	1
		100.0%	0.0%	0.0%	50.0%	50.0%	0.0%	50.0%	50.0%	0.0%	0.0%	50.0%	50.0%
	Total	12	0	0	11	1	4	5	2	0	1	1	11
		100.0%	0.0%	0.0%	91.7%	8.3%	33.3%	41.7%	16.7%	0.0%	8.3%	8.3%	91.7%

References

- Adger, C., & Christian, D. (2007). *Sociolinguistic variation and education*. In R. Bayley & C. Lucas (Eds.), *Sociolinguistic variation: Theories, methods, and applications* (pp. 237-253). Cambridge: Cambridge University Press.
- Adger, C., Wolfram, M. and Detwyler, J., & Harry, B. (1993). Confronting dialect minority issues in special education: Reactive and proactive perspectives. In *Proceedings of the Third National Research Symposium on Limited English Proficient Student Issues: Focus on middle and high school issues* (Vol. 2, pp. 737-762). Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs.
- Akhlaghi, F. (2020). Iran Linguistic Atlas project (ILA): Past and present. *Iranian Studies*, 53(3-4), 623-659.
- Albury, N. (2014). Introducing the folk linguistics of language policy. *International Journal of Language Studies*, 8(3), 85-106.
- Albury, N. (2017). How folk linguistic methods can support critical sociolinguistics. *Lingua*, 199, 36-49.
- Aliakbari, M., Gheitasi, M., & Anonby, E. (2015). On language distribution in Ilam province, Iran. *Iranian Studies*, 48(6), 835-850.
- Allen, I. E., & Seaman, C. A. (2007). Likert scales and data analyses. *Quality Progress*, 40(7), 64-65.
- Amery, R. (2009). Phoenix or relic? Documentation of languages with revitalization in mind. *Language Documentation and Conservation*, 3(2), 138-148.

- Anonby, E., Taheri-Ardali, M. et al. (2015-2021). *Atlas of the languages of Iran*. Ottawa: Geomatics and Cartographic Research Centre. Available at: <http://iranatlas.net>.
- Anonby, E., Hayes, A., & Oikle, R. (2020). A multi-dimensional approach to classification of Iran's languages. In R.K. Larson, S. Moradi & V. Samiian (Eds.), *Advances in Iranian linguistics* (Vol. 351, pp. 29-56). Amsterdam: John Benjamins Publishing Company.
- Anonby, E., & Sabethematabadi, P. (2019). Representing complementary user perspectives in a language atlas. In D. R. Fraser Taylor, E. Anonby & K. Murasugi (Eds.), *Further developments in the theory and practice of cybercartography: International dimensions and language mapping* (pp. 413-440). Amsterdam: Elsevier.
- Anonby, E., Sabethematabadi, P., & Hayes, A. (2016, February). *Reconciling contradictory perspectives of language identity in the Atlas of the Languages of Iran* [Paper presentation]. Cognitive Science Colloquium Series, Carleton University, Canada.
- Anonby, E., Taheri-Ardali, M., & Hayes, A. (2019). The Atlas of the Languages of Iran (ALI): A research overview. *Iranian Studies*, 52(1-2), 199-230.
<https://doi.org/10.1080/00210862.2019.1573135>
- Anonby, E., & Yousefian, P. (2011). *Adaptive multilinguals a survey of language on Larak Island*. Uppsala: Acta Universitatis Upsaliensis.
- Armstrong, R. L. (1987). The midpoint on a five-point Likert-type scale. *Perceptual and Motor Skills*, 64(2), 359-362.
- Asgharzadeh, A. (2007). *Iran and the challenge of diversity: Islamic fundamentalism, Aryanist racism, and democratic struggles* (1st ed.). New York: Palgrave Macmillan.

- Bani-Shoraka, H. (2005). The Iranian language policy of the twentieth century: The case of Azerbaijani in Tehran. In A. Rabo & B. Utas (Eds.), *The role of the state in West Asia* (pp. 141-150). Istanbul: Swedish Research Institute in Istanbul.
- Barbour, R. S. (1998). Mixing qualitative methods: Quality assurance or qualitative quagmire? *Qualitative Health Research*, 8(3), 352-361.
- Bateni, M. R. (2014). *Masāel-e zabānshenāsi-e novin* [New linguistic issues]. Tehran: Āgāh.
- Benson, E. J. (2003). Folk linguistic perceptions and the mapping of dialect boundaries. *American Speech*, 78(3), 307-330.
- Bhatia, V. K., Rodney, H. J., & Flowerdew, J. (2008). *Advances in discourse studies*. New York: Routledge.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston.
- Bloomfield, L. (1944). Secondary and tertiary responses to language. *Language (Baltimore)*, 20(2), 45-55.
- Bolarinwa, O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, 22(4), 195-201.
- Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, 50(2).
Online at: https://archives.joe.org/joe/2012april/pdf/JOE_v50_2tt2.pdf
- Borjian, H. (2008). The extinct language of Gurgān: Its sources and origins. *Journal of the American Oriental Society*, 128(4), 681-707.

- Borjian, H. (2013). Farvi dialect. In *Encyclopædia Iranica*. New York: Center for Iranian Studies, Columbia University. Online at: <https://iranicaonline.org/articles/farvi-dialect>
- Boroujerdi, M. (1998). Contesting nationalist constructions of Iranian identity. *Critique: Critical Middle Eastern Studies*, 7(12), 43-55.
- Bruk, S. I., & Apenchenko, V. S. (Eds.). (1964). *Atlas narodov mira* [Atlas of the World's Peoples]. Moscow: Glavnoye Upravleniye Geodezii i Kartografii.
- Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of mixed methods research*, 1(1), 8-22.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge.
- Cahill, M. C., & Karan, E. (2008). Factors in designing effective orthographies for unwritten languages. *SIL Electronic Working Papers 2008(001)*, 1-15. Dallas, Texas: Michael Cahill, Elke Karan, and SIL International.
- Cameron, D. (2014). Gender and language ideologies. In S. Ehrlich, M. Meyerhoff & J. Holmes (Eds.), *The handbook of language, gender, and sexuality* (2nd ed., pp. 281-296). John Wiley & Sons, Inc.
- Carifio, L., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42, 1150-1152.
- Chomsky, N. (1972). *Language and mind* (2nd ed.). New York: Harcourt Brace Jovanovich.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, Evanston and London: Harper & Row.

- Chambers, J., & Trudgill, P. (1998). *Dialectology* (2nd ed.). Cambridge: Cambridge University Press.
- Clason, D. L., & Dormody, T. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education*, 35(4), 31-35.
- Cook, G. (2003). *Applied linguistics*. Oxford: Oxford University Press.
- Cortina, J. M. (1993). What is coefficient alpha?: An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104.
- Coupland, N. (1988). *Dialect in use: Sociolinguistic variation in Cardiff English*. Cardiff: University of Wales Press.
- Coupland, N., Williams, A., & Garrett, P. (1999). “Welshness” and “Englishness” as attitudinal dimensions of English language varieties in Wales. In D. R. Preston (Ed.), *Handbook of perceptual dialectology: Volume 1* (pp. 333-343). Amsterdam: John Benjamins Publishing Company.
- Cramer, J. (2016a). *Perceptual dialectology*. (Oxford handbooks online. Linguistics). New York: Oxford University Press. Online at:
<https://doi.org/10.1093/oxfordhb/9780199935345.013.60>
- Cramer, J. (2016b). Rural vs. urban: Perception and production of identity in a border city. In J. Cramer & C. Montgomery (Eds.), *Cityscapes and perceptual dialectology: Global perspectives on non-linguists' knowledge of the dialect landscape (Language and social life)* (Vol. 5, pp. 27-53). Boston / Berlin: De Gruyter.
- Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Los Angeles: Sage Publications.

- Dabir-Moghaddam, M. (2008). Language, dialect, accent, and variety: Their use in the writings of Iranian and non-Iranian scholars. *Journal of Adab Pazhuhi*, 2(5), 91-128.
- Dailey O’Cain (1997/1999). The perception of post-unification German regional speech. In D. R. Preston (Ed.), *Handbook of perceptual dialectology: Volume 1* (pp. 227-242). Amsterdam / Philadelphia: John Benjamins Publishing Company. (Original work published 1997).
- Daniels, P. T., & Bright, W. (1996). *The world’s writing systems*. Oxford: Oxford University Press.
- Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*. Chicago: Aldine.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). New York: Routledge.
- Dziuban, C.D., & Shirkey, E.C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, 31, 358-361.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2021). *Ethnologue: Languages of the World, 24th edition*. Dallas: SIL International. Online at: <http://www.ethnologue.com>.
- Edwards, J. (2009). *Language and identity: An introduction*. Cambridge: Cambridge University Press.
- Eira, C. (1998). Authority and discourse: Towards a model of orthography selection. *Written Language and Literacy*, 1(2), 171-224.

- Elling, R. (2013). *Minorities in Iran: Nationalism and ethnicity after Khomeini* (1st ed.). New York: Palgrave Macmillan.
- Eppler, E., & Benedikt, J. (2017). A perceptual dialectological approach to linguistic variation and spatial analysis of Kurdish varieties. *Journal of Linguistic Geography*, 5(2), 109-130.
- Erzerberger, C., & Prein, G. (1997). Triangulation: Validity and empirically based hypothesis construction. *Quality and Quantity*, 31, 141-154.
- Evans, B. E. (2016). City talk and Country talk: Perceptions of urban and rural English in Washington state. In J. Cramer & C. Montgomery (Eds.), *Cityscapes and perceptual dialectology: Global perspectives on non-linguists' knowledge of the dialect landscape (Language and social life)* (Vol. 5, pp. 55-72). Boston / Berlin: De Gruyter.
- Farmer, T., Robinson, K., Elliott, S. J., & Eyles, J. (2006). Developing and implementing a triangulation protocol for qualitative health research. *Qualitative Health Research*, 16(3), 377-394.
- Fielding, N. G., & Fielding, J. L. (1986). *Linking data*. CA: Sage Publications.
- Fishman, J. A. (1965). Who speaks what language to whom and when? *La Linguistique*, 1(2), 67-88.
- Fishman, J. A. (1971). The links between micro- and macro-sociolinguistics in the study of who speaks what language to whom and when. In J. A. Fishman, R. L. Cooper & R. M. Newman (Eds.), *Bilingualism in the barrio* (pp. 583-604). Bloomington: Indiana University Press.
- Fishman, J. A. (1972). *Language and nationalism: Two integrative essays*. Rowley, Mass: Newbury.

- Fishman, J. A. (1977). Preface: Advances in the creation and revision of writing systems. In J. A. Fishman (Ed.), *Advances in the creation and revision of writing systems* (pp. XI-XXVIII). Berlin / Boston: De Gruyter Mouton.
- Fridlund, B. (2010). The dissertation book; Should it be a monograph or a compilation thesis? *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, 9(3), 144-145.
- Geerts, G., Van Den Broeck, J., & Verdoodt, A. (1977). Successes and failures in Dutch spelling reform. In J. A. Fishman (Ed.), *Advances in the creation and revision of writing systems* (pp. 179-245). Berlin / Boston: De Gruyter Mouton.
- Gooskens, C. (2007). The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development*, 28(6), 445-467.
- Gooskens, C., & Schneider, C. (2016). Testing mutual intelligibility between closely related languages in an oral society. *Language Documentation and Conservation*, 10, 278-305.
- Gooskens, C., van Heuven, V. J., Golubović, J., Schüppert, A., Swarte, F., & Voigt, S. (2018). Mutual intelligibility between closely related languages in Europe. *International Journal of Multilingualism*, 15(2), 169-193.
- Grinevald, C., & Sinha C. (2016). North-South relations in linguistic science: Collaboration or colonialism? In L. Filipović & M. Pütz (Eds.), *Endangered languages and languages in danger: Issues of documentation, policy, and language rights* (pp. 25-43). (IMPACT: Studies in language and society 42). Amsterdam / Philadelphia, PA: John Benjamins Publishing Company.

- Grootaers, W.A. (1999). The discussion surrounding the subjective boundaries of dialects. In D. R. Preston (Ed.) & L. Kuiper (Trans.), *Handbook of perceptual dialectology: Volume 1* (pp. 115-130). Amsterdam / Philadelphia: John Benjamins Publishing Company. (Original work published 1964).
- Gumperz, J. J. (1964). Linguistic and social interaction in two communities. *American Anthropologist*, 66(6), 137-153.
- Gumperz, J. J. (1970). *Sociolinguistics and communication in small groups*. Berkeley: University of California.
- Gumperz, J. J., & Hymes, D. H. (1972). *Directions in sociolinguistics: The ethnography of communication*. New York: Holt, Rinehart and Winston.
- Haig, G., & Öpengin, E. (2014). Introduction to special Issue - Kurdish: A critical research overview. *Kurdish Studies*, 2(2), 99-122.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2021). *Glottolog* 4.4. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online at: <http://glottolog.org>.
- Hartley, L. (1999). A view from the West: Perceptions of U.S. dialects by Oregon residents. In D. R. Preston (Ed.), *Handbook of perceptual dialectology: Volume 1* (pp. 315-332). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Hassanpour, A. (1991). State policy on the Kurdish language: The politics of status planning. *Kurdish Times*, 4(1), 42-85.

- Hassanpour, A. (1992). *Nationalism and language in Kurdistan, 1918-1985*. San Francisco: Mellen Research University Press.
- Hassanpour, A. (1993). Communications. *Middle East Journal*, 47(3), 571-577.
- Hassanpour, A. (1994). The Kurdish experience. *Middle East Report*, 24(189), 2-7, 23.
- Hassanpour, A. (1998). The identity of Hewrami speakers: Reflections on the theory and ideology of comparative philology. In A. Soltani (Ed.), *Anthology of Gorani Kurdish poetry, compiled by A. M. Mardoukhi (1739–1797)* (pp. 35-49). London: Soane Trust for Kurdistan.
- Hassanpour, A. (2000). The politics of a-political linguistics: Linguists and linguisticide. In R. Phillipson (Ed.), *Rights to language: Equity, power and education* (pp. 33-39). New Jersey: Lawrence Erlbaum Associates.
- Hassanpour, A. (2012). The indivisibility of the nation and its linguistic divisions. *International Journal of the Sociology of Language*, 2012(217), 49-73.
- Haugen, E. (1966). Dialect, language, nation. *American Anthropologist*, 68(4), 922-935.
- Heeringa, W., & Nerbonne, J. (2001). Dialect areas and dialect continua. *Language Variation and Change*, 13(3), 375-400.
- Hemming, P. J., & Madge, N. (2011). Researching children, youth and religion: Identity, complexity and agency. *Childhood*, 19(1), 38-51.
- Hill, J. H. (2002). “Expert rhetorics” in advocacy for endangered languages: Who is listening, and what do they hear? *Journal of Linguistic Anthropology*, 12(2), 119-133.
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104-121.

- Hoenigswald, H. (1966). A proposal for the study of folk-linguistics. In W. Bright (Ed.), *Sociolinguistics* (pp. 16-26). The Hague: Mouton.
- Hourcade, B., Lebugle-Mojdehi, A., Giraudet, E., Nejad, P. R., & Atanasiu, V. (2012). *Irancarto*. Paris and Tehran: CNRS / University of Tehran. Online at: <http://www.irancarto.cnrs.fr/index.php?l=en>
- Hudson, R. A. (2001). *Sociolinguistics* (2nd ed.). Cambridge: Cambridge University Press.
- Hunt, S. D. (1991). *Modern marketing theory*. Cincinnati, OH: South-Western.
- Hymes, D. H. (1964). Introduction: Toward ethnographies of communication. *American Anthropologist*, 66(6), 1-34.
- Hymes, D. H. (1967). Models of interaction of language and social setting. *Journal of Social Issues*, 33(2), 8-28.
- Hymes, D. H. (1972). Models of the interaction of language and social life. In J. J. Gumperz & D. H. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 35-71). New York: Holt, Rinehart and Winston.
- Hymes, D. H. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.
- Hymes, D. H. (1989). Ways of speaking. In R. Bauman & J. Sherzer (Eds.), *Explorations in the ethnography of speaking* (2nd ed., pp. 433-451). Cambridge: Cambridge University Press.
- Inoue, F. (1996). Subjective dialect division in Great Britain. *American Speech*, 71(2), 142-161.
- Izady, M. M. (2014). Iran: Linguistic composition. Online at: https://gulf2000.columbia.edu/images/maps/Iran_Languages_2014_lg.png

- Jahani, C. (2001). Balochi. In J. Garry & C. Rubino (Eds.), *Facts about the world's languages* (pp. 59-64). New York: H.W. Wilson.
- Jahani, C. (2005). State control and its impact on language in Balochistan. In A. Rabo & B. Utas (Eds.), *The role of the state in West Asia* (pp. 151-163). Istanbul: Swedish Research Institute in Istanbul.
- Jahani, C., & Korn, A. (2009). Balochi. In G. Windfuhr (Eds.), *The Iranian languages* (pp. 634-692). Oxon / New York: Routledge.
- Jaffe, A. (2000). Introduction: Non-standard orthography and non-standard speech. *Journal of Sociolinguistics*, 4(4), 497-513.
- Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38(12), 1217-1218.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24, 602-611.
- Joseph, J. E. (2004). *Language and identity: National, ethnic, religious*. London: Palgrave Macmillan.
- Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, 34(1), 111-117.
- Karan, E. (2006). *Writing system development and reform: A process* [Unpublished M.A. thesis]. University of North Dakota. Online at:
https://arts-sciences.und.edu/academics/summer-institute-of-linguistics/theses/_files/docs/2006-karan-elke.pdf

- Kia, S. (1961). *Rāhnamāy-e gerdāvāri-e guyeshha* [Guide to collecting dialects]. Tehran: Enteshārāt-e edārehy-e farhang-e āmme.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: The Guilford Press.
- Kremer, L. (1999). The Dutch-German national border as a subjective dialect boundary. In D. R. Preston (Ed.) & J. Dailey-O’Cain (Trans.), *Handbook of perceptual dialectology: Volume 1* (pp. 31-36). Amsterdam / Philadelphia: John Benjamins Publishing Company. (Original work published 1984).
- Kuiper, L. (1999). Variation and the norm: Parisian perceptions of regional French. In D. R. Preston (Ed.), *Handbook of perceptual dialectology: Volume 1* (pp. 243-262). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, Calif: Sage Publications.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lavandera, B. R. (1988). The study of language in its socio-cultural context. *Linguistics: The Cambridge survey*, 4, 1-13.
- Le Page, R. B. (1968). Problems of Description in Multilingual communities. *Transactions of the Philological Society*, 67(1), 189-212.
- Leonard, W. Y., & Haynes, E. (2010). Making “collaboration” collaborative: An examination of perspectives that frame linguistic field research. *Language Documentation and Conservation*, 4, 269-293.

- Lichter, D. T., & Brown, D. L. (2011). Rural America in an urban society: Changing spatial and social boundaries. *Annual Review of Sociology*, 37, 565-592.
- Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60(1), 10-13.
- Lomax, R.G., & Hahs-Vaughn, D. L. (2012). *An introduction to statistical concepts* (3rd ed.). New York: Routledge.
- Long, D. (1999). Mapping nonlinguists' evaluations of Japanese language variation. In D. R. Preston (Ed.), *Handbook of perceptual dialectology: Volume 1* (pp. 199-226). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Makoni, S., & Pennycook, A. (2007). *Disinventing and reconstituting languages*. Clevedon: Multilingual Matters Ltd.
- Mariou, E. (2017). My language, my mother tongue: Competing language ideologies and linguistic diversity among speakers of standard and non-standard varieties. *International Journal of Bilingual Education and Bilingualism*, 20(1), 20-14.
- Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice*, 13(6), 522-526.
- May, S. (2006). Language policy and minority rights. In T. Ricento (Eds.), *An introduction to language policy: Theory and method* (pp. 255-272). Malden, MA: Blackwell Publishing Ltd.
- May, S. (2012). *Language and minority rights: Ethnicity, nationalism and the politics of language*. London / New York: Routledge.
- McCarus, E. N. (2009). Kurdish. In G. Windfuhr (Ed.), *The Iranian languages* (pp. 587-633). Oxon / New York: Routledge.

- Medsker, G., Williams, L., & Holahan, P. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*, 20(2), 439-464.
- Miller, C., Livingston, J., Vinson, M., & Triebwasser Prado, T. (2006). *Persian dialects as spoken in Iran*. Maryland: University of Maryland center.
- Mirshahidi, S. (2017). I find you attractive but I don't trust you: The case of language attitudes in Iran. *Journal of Multilingual and Multicultural Development*, 38(2), 146-159.
- Mirvahedi, S. H. (2017). Exploring family language policies among Azerbaijani-speaking families in the city of Tabriz, Iran. In J. Macalister & S. H. Mirvahedi (Eds.), *Family language policies in a multilingual world: Opportunities, challenges, and consequences* (pp. 74-95). London: Routledge.
- Mithun, M. (2007). What is a language? Documentation for diverse and evolving audiences. *Sprachtypologie und Universalienforschung: STUF*, 60(1), 42-55.
- Modarresi, Y. (2014). *An introduction to sociolinguistics* (4th ed.). Tehran: Institute for Humanities and Cultural Studies.
- Mohammadbakhsh, B., Fathiazar, E., Hobbi, A., & Ghodrathpour, M. (2012). Globalization and local and global identities among Iranian students. *International Journal of Intercultural Relations*, 36(1), 14-21.
- Montgomery, C. (2007). *Northern English dialects: A Perceptual Approach* [Unpublished doctoral dissertation]. University of Sheffield. Online at:
<http://etheses.whiterose.ac.uk/1203/>

- Moradi, S. (2019). Languages of Iran: Overview and critical assessment. In S. D. Brunn & R. Kehrein (Eds.), *Handbook of the changing world language map* (pp. 1171-1202). Cham: Springer International Publishing.
- Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing research*, 40(2), 120-123.
- Murphy, A. B. (1998). European languages. In T. Unwin (Ed.), *A European geography* (pp. 34-50). Harlow / New York: Longman.
- Naghzguy-Kohan, M. (2016). *Zabān-e Farsi dar gozar-e zamān (research articles)* [Persian (Farsi) language over time]. Tehran: Chāpār Nashr.
- Nettle, D. (1999). Is the rate of linguistic change constant?. *Lingua*, 108(2), 119-136.
- Neuman, W., & Robson, K. (2012). *Basics of social research: Qualitative and quantitative approaches* (2nd Canadian ed.). Toronto: Pearson Canada.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 62-85.
- Niedzielski, N., & Preston, D. R. (2000). *Folk linguistics*. Berlin: Mouton De Gruyter.
- Nomoto, K. (1999). Consciousness of linguistic boundaries and actual linguistic boundaries. In D. R. Preston (Ed.) & D. Long (Trans.), *Handbook of perceptual dialectology: Volume 1* (pp. 63-69). Amsterdam: John Benjamins Publishing company. (Original work published 1963).
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education: Theory and Practice*, 15(5), 625-632.
- Nunnally, J. C. (1976). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

- Ochs, E., & Schieffelin, B. (2006). The impact of language socialization on grammatical development. In C. Jourdan & K. Tuite (Eds.), *Language, culture and society* (pp. 168-189). Cambridge: Cambridge University Press.
- Odango, E. L. (2015). May sasabihin ang kabataan ‘the youth have something to say’: Youth perspectives on language shift and linguistic identity. *Language Documentation & Conservation*, 9, 32-58.
- Orywal, E. (Eds.). (1988). Karte 10, Vorderer Orient: Sprachen und Dialekte [Map 10, Middle East: languages and dialects]. *Tübinger Atlas des Vorderen Orients (TAVO)*, series A, vol. 8. Wiesbaden: Reichert.
- Pasquale, M. D., & Preston, D. R. (2013). The folk linguistics of language teaching and learning. In K. Drozdziel-Szelest & M. Pawlak (Eds.), *Psycholinguistic and Sociolinguistic Perspectives on Second Language Learning and Teaching* (1st ed., pp. 163-174). Heidelberg: Springer Berlin Heidelberg.
- Paveau, M. (2011). Do non-linguists practice linguistics?: An anti-eliminative approach to folk theories. *AILA Review*, 24, 40-54.
- Perry, J. R. (1998). Languages and dialects: Islamic period. *Iranian Studies*, 31(3-4), 517-525.
- Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences* (6th ed.). New York: Routledge.
- Preston, D. R. (1982). Perceptual dialectology: Mental maps of United States dialects from a Hawaiian perspective. *Hawaii Working Papers in Linguistics*, 14(2), 5-49.
- Preston, D. R. (1989). *Perceptual dialectology: Nonlinguists' views of areal linguistics*. Dordrecht: Foris Publications Holland.

- Preston, D. R. (1993). The uses of folk linguistics. *International Journal of Applied Linguistics*, 3(2), 181-259.
- Preston, D. R. (1994). Content-oriented discourse analysis and folk linguistics. *Language Sciences*, 16(2), 285-331.
- Preston, D. R. (Ed.). (1999a). *Handbook of perceptual dialectology: Volume 1*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Preston, D. R. (1999b). A language attitude approach to the perception of regional variety. In D. R. Preston (Ed.), *A handbook of perceptual dialectology: Volume 1* (pp. 359-373). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Preston, D. R. (1999c). A language attitude analysis of regional U.S. speech: Is Northern U.S. English not friendly enough? *Cuadernos de Filología Inglesa*, 8, 129-146.
- Preston, D. R. (2005). What is folk linguistics? Why should you care? *Lingua Posnaniensis*, 47, 143-162.
- Preston, D. R. (2011). Methods in (applied) folk linguistics: Getting into the minds of the folk. *AILA Review*, 24, 15-39.
- Preston, D. R. (2013). Linguistic insecurity forty years later. *Journal of English Linguistics*, 41(4), 304-331.
- Preston, D. R. (2016, July 19-22). *Why folk linguists should hate language awareness* [Paper presentation]. Association for language awareness, Vienna university of economics and Business, Austria.
- Priestly, T. (1992). Problems in the creation of an orthography: Functional load, interference, and political preferences. *Slavic and East European Journal*, 36, 302-316.

- Redard, G. (1974). État des travaux et publication: quelques cartes onomasiologiques [The state of work and publication: Some onomasiological maps]. In G. Redard et al. (Eds.), *Arbeitspapiere* 13 [Working papers 13]: *Atlas des parlers Iraniens-Atlas de l'Afghanistan* [Atlas of the Iranian languages- Atlas of Afghanistan] (pp. 7-19). Bern: Universität Bern, Institut für Sprachwissenschaft.
- Rensink, W. G. (1999). Informant classification of dialects. In D. R. Preston (Ed.) & D. R. Preston (Trans.), *Handbook of perceptual dialectology: Volume 1* (pp. 3-7). Amsterdam / Philadelphia: John Benjamins Publishing Company. (Original work published 1955).
- Rezaei, S., & Bahrami, A. (2019). Attitudes toward Kurdish in the city of Ilam in Iran. In S. H. Mirvahedi (Ed.), *The sociolinguistics of Iran's languages at home and abroad: The case of Persian, Azerbaijani, and Kurdish* (pp. 77-106). Cham, Switzerland: Palgrave Macmillan.
- Rezaei, S., & Farnia, M. (2016, June). *Language attitude and identity among Armenians in diaspora: A case of Armenians in the city of Isfahan* [Paper presentation]. Sociolinguistics Symposium 21, Universidad de Murcia, Murcia, Spain.
- Rezaei, S., Khatib, M., & Baleghizadeh, S. (2014). Language identity among Iranian English language learners: A nationwide survey. *Journal of Multilingual and Multicultural Development*, 35(5), 527-536.
- Rezaei, S., Latifi, A., & Nematzadeh, A. (2017). Attitude towards Azeri language in Iran: A large-scale survey research. *Journal of Multilingual and Multicultural Development*, 38(10), 931-941.

- Rezaei, S., & Tadayyon, M. (2018). Linguistic landscape in the city of Isfahan in Iran: The representation of languages and identities in Julfa. *Multilingua*, 37(6), 701-720.
- Robinson, C., & Gadellii, K. (2003). Writing unwritten languages: A guide to the process. UNESCO. Online at:
http://portal.unesco.org/education/en/ev.phpurl_id=30738&url_do=do_printpage&url_section=201.html
- Rubin, H. J., & Rubin, I. (1995). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage Publications.
- Rust, V. D. (2012). Minority education in Iran. In J. Banks & Gale Group (Eds.), *Encyclopedia of diversity in education* (A Sage Reference Publication). Thousand Oaks, CA: Sage Publications.
- Ryan, E. B., Giles, H., & Sebastian, R. J. (1982). An integrative perspective for the study of language attitudes toward language variation. In E. B. Ryan & H. Giles (Ed.), *Attitudes towards language variation: Social and applied contexts* (pp. 1-19). London: Edward Arnold.
- Sabthematabadi, P., & Anonby, E. (forthcoming). An analysis of folk linguistic perspectives on language identification in Iran. *The International Association of Applied Linguistics (AILA)*.
- Sabthematabadi, P., & Anonby, E. (forthcoming). Language identification in Iran: A sociolinguistic model and questionnaire. *Language Related Research*.

- Sabethematabadi, P. & Anonby, E. (2021). *Languages of Iran: Language identification questionnaire*. Toronto: Scholars Portal. Online at:
<https://doi.org/10.5683/SP2/DEZXBI>.
- Sabethematabadi, P., Anonby, E., & Sheyholislami, J. (2017, June). *A comparative analysis of language classification taxonomies in Tehrani, Kurdish, and Abade'i language communities* [Paper presentation]. The 3rd International Conference on Kurdish Studies (ICKS3), Exeter University, Exeter, UK.
- Sadeghi, A. A. (1970). Zabān-e Farsi va gunehhāy-e mokhtalef-e ān [Persian language and its varieties]. *Farhang va Zendegi*, 2, 61-66.
- Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). London: Sage Publications.
- Saleh, A. (2013). *Ethnic identity and the state in Iran* (1st ed.). New York: Palgrave Macmillan.
- Saville-Troike, M. (2003). *The ethnography of communication: An introduction* (3rd ed.). Malden, MA: Blackwell Publishing Ltd.
- Schiffman, H. (2006). Linguistic policy and linguistic culture. In T. Ricento (Ed.), *An introduction to language policy: Theory and method* (pp. 111-125). Malden, MA: Blackwell Publishing Ltd.
- Schmitt, R. (1989). *Compendium linguarum iranicarum* [Compendium of Iranian languages]. Wiesbaden: Reichart.
- Schoenfeldt, L. F. (1984). Psychometric properties of organizational research instruments. In T. S. Bateman & G. R. Ferris (Eds.), *Method and analysis in organizational research* (pp. 68-80). Reston, VA: Reston.

- Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to structural equation modeling* (4th ed.). New York: Routledge.
- Sebba, M. (1998). Phonology meets ideology: The meaning of orthographic practices in British Creole. *Language Problems and Language Planning*, 22(1), 19-47.
- Sebba, M. (2009). Sociolinguistic approaches to writing systems research. *Writing Systems Research*, 1(1), 35-49.
- Shahiditabar, M., & Pourghasemian, H. (2016). A cross-cultural view of Iranian multilingualism: The case of Azerbaijani Turkish, Persian, Arabic and Kurdish contact. *Global Journal of Social Sciences Studies*, 2(2), 72-82.
- Shaw, E. (1999). A guide to the qualitative research process: Evidence from a small firm study. *Qualitative Market Research: An International Journal*, 2(2), 59-70.
- Sheyholislami, J. (2012). Kurdish in Iran: A case of restricted and controlled tolerance. *International Journal of the Sociology of Language*, 217, 19-47.
- Sheyholislami, J. (2015). Language varieties of the Kurds. In W. Taucher, M. Vogl & P. Webinger (Eds.), *The Kurds: History, religion, language, politics* (pp. 30-51). Vienna, Austria: Austrian Federal Ministry of the Interior.
- Sheyholislami, J., & Sharifi, A. (2016). "It is the hardest to keep": Kurdish as a heritage language in the United States. *International Journal of the Sociology of Language*, 237, 75-98.
- Sheyholislami, J. (2019). Language as a problem: Language policy and rights in Kurdistan-Iran. In S. Akin, J. Blau, F. Hellot, S. Mousset, H. Mede, S. Alsancakli & A. Ahmed (Eds.), *Etudes Kurde* (pp. 99-134). Paris, France: The Kurdish institute of Paris.

- Shih, F. J. (1998). Triangulation in nursing research: Issues of conceptual clarity and purpose. *Journal of Advanced Nursing*, 28(3), 631-641.
- Shultz, K. S., Whitney, D. J., & Zickar, M. J. (2014). *Measurement theory in action: Case studies and exercises* (2nd ed.). London / New York: Routledge.
- Sibata, T. (1999). Consciousness of dialect boundaries. In D. R. Preston (Ed.) & D. Long (Trans.), *Handbook of perceptual dialectology: Volume 1* (pp. 39-62). Amsterdam: John Benjamins Publishing Company. (Original work published 1959).
- Skutnabb-Kangas, T., Sheyholislami, J., & Hassanpour, A. (2012). Concluding remarks. *International Journal of the Sociology of Language*, 2012(217), 181-187.
- Stilo, D. (1981). The Tati language group in the sociolinguistic context of Northwestern Iran and transcaucasia. *Iranian Studies*, 14(3/4), 137-187.
- Stilo, D. (2007a). Isfahān, xxi. Provincial dialects. *Encyclopædia Iranica*, 14(1), 93-112. New York: Center for Iranian Studies, Columbia University. Online at: <http://www.iranicaonline.org/articles/isfahan-xxi-provincial-dialects>
- Stilo, D. (2007b). *An introduction to the atlas of the Araxes-Iran (Outer Fertile Crescent) Linguistic Area* [Paper presentation]. The School of Oriental and African Studies, London, UK.
- Stilo, D. (2016, August). *The Araxes-Iran linguistic area with a concentration on Kurdoophone varieties* [Paper presentation]. The 3rd International Conference on Kurdish Linguistics (ICKL3), University of Amsterdam, Amsterdam, the Netherlands.

- Sturgis, P., Roberts, C., & Smith, P. (2014). Middle alternatives revisited: How the neither/nor response acts as a way of saying “I don’t know”? *Sociological Methods & Research*, 43(1), 15-38.
- Souza, A. (2016). Language and religious identities. In S. Preece (Ed.), *The Routledge handbook of language and identity* (pp. 195-209). London / New York: Routledge.
- Sullivan, G. M., & Artino, J. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541-542.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Upper Saddle River, NJ: Pearson.
- Taheri-Ardali, M., Anonby, E., & Hayes, A. (2021). *Atlas-e barkhatt-e zabānhā-ye Iran: Tarrāhi, raveshshenāsi va natāyej-e avvaliyyeh* [The online *Atlas of the Languages of Iran: Design, methodology and initial results*]. *Language Related Research*, 12(2), 231-291.
- Thomas, D. R. (2006). A General inductive approach for analyzing qualitative evaluation data. *The American Journal of Evaluation*, 27(2), 237-246.
- Tohidi, N. (2009). Ethnicity and religious minority politics in Iran. In A. Gheissari (Ed.), *Contemporary Iran: Economy, society, politics* (pp. 299-323). Oxford: Oxford University Press.
- Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

- Wardhaugh, R. (2006). *An introduction to sociolinguistics (5th ed.)*. Cambridge: Blackwell Publishing Ltd.
- Weijnen, A. A. (1999). On the value of subjective dialect boundaries. In D. R. Preston (Ed.) & J. Dailey-O'Cain (Trans.), *Handbook of perceptual dialectology: Volume 1* (pp. 131-134). Amsterdam / Philadelphia: John Benjamins Publishing Company. (Original work published 1946).
- Whittaker, T. A. (2016). Structural equation modeling. In K. A. Pituch & J. P. Stevens (Eds.), *Applied multivariate statistics for the social sciences (6th ed., pp. 639-746)*. New York: Routledge.
- Williams, A., Garrett, P., & Coupland, N. (1999). Dialect recognition. In D. R. Preston (Ed.), *Handbook of perceptual dialectology: Volume 1* (pp. 345-358). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Williams, P., Stevenson, I., Nicholas, D., Watkinson, A., & Rowlands, I. (2009). The role and future of the monograph in arts and humanities research. *Aslib Proceedings*, 61(1), 67-82.
- Wilton, A., & Stegu, M. (2011). Bringing the 'folk' into applied linguistics: An introduction. *AILA Review*, 24, 1-14.
- Windfuhr, G. (1995). Dialectology. *Encyclopædia Iranica*, 7(4), 362-370. New York: Center for Iranian Studies, Columbia University.
- Windfuhr, G. (2006). Iran, vii. non-Iranian languages, 9. Arabic. *Encyclopædia Iranica*, 13(4), 401-404. New York: Center for Iranian Studies, Columbia University.

- Windfuhr, G. (2009). Dialectology and topics. In G. Windfuhr (Ed.), *The Iranian languages* (pp. 5-42). Oxon / New York: Routledge.
- Windfuhr, G. (2012). "Iran vii. Non-Iranian languages (Turkic languages)." *Encyclopaedia Iranica*. Online at:
<http://www.iranicaonline.org/articles/iran-vii7-turkic-languages>
- Wolfram, W. (2017) Dialect in society. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics*. Blackwell Publishing Ltd. Online at:
http://liduaeka.weebly.com/uploads/1/0/7/6/10761275/7_dialect_in_society.pdf
- Wolfram, W., & Schilling, N. (2016). *American English: Dialects and variation* (3rd ed.). John Wiley & Sons, Inc.
- Yaeger-Dror, M. (2014). Religion as a sociolinguistic variable: When religious identity shapes linguistic patterns. *Language and Linguistics Compass*, 8(11), 577-589.