

Georgia State University

ScholarWorks @ Georgia State University

Applied Linguistics and English as a Second
Language Dissertations

Department of Applied Linguistics and English
as a Second Language

12-10-2018

Discerning the Language Assessment Literacy of EFL Teachers in Uzbekistan: An Individual, Social, and Sociohistorical Teacher Cognition Inquiry

David Lawrence Chiesa

Follow this and additional works at: https://scholarworks.gsu.edu/alesl_diss

Recommended Citation

Chiesa, David Lawrence, "Discerning the Language Assessment Literacy of EFL Teachers in Uzbekistan: An Individual, Social, and Sociohistorical Teacher Cognition Inquiry." Dissertation, Georgia State University, 2018.

https://scholarworks.gsu.edu/alesl_diss/49

This Dissertation is brought to you for free and open access by the Department of Applied Linguistics and English as a Second Language at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Applied Linguistics and English as a Second Language Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

DISCERNING THE LANGUAGE ASSESSMENT LITERACY OF EFL TEACHERS IN
UZBEKISTAN: AN INDIVIDUAL, SOCIAL, AND SOCIOHISTORICAL TEACHER
COGNITION INQUIRY

by

DAVID CHIESA

Under the Direction of Co-Chairs, Sara Cushing, Ph.D. & John Murphy, Ed.D.

ABSTRACT

Language teacher cognitions can be complex, ranging over many different subjects; they can be dynamic, changing over time and under different influences; and they can be systematic, forming unified and cohesive personal and practical theories (Feryok, 2010). One subject matter area that has not been explored in the literature is the relationship between second language (L2) teacher cognition and assessment, also known as (language) assessment literacy – the level of a teacher's engagement with constructing, using, and interpreting a variety of assessment procedures to make decisions about a learner's language ability (Taylor, 2013). I examine L2 teachers'

cognitions about assessment at the individual level, and then analyze how micro-institutional (social) and macro-sociocultural aspects of their lives as language teachers (including past, present, and future aspects) are shaping teachers' assessment practices.

This investigation focused on a group of 96 in-service university English language teachers in Uzbekistan. The three overarching research questions are: (1) To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy? (2) How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students? (3) What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students? The data were collected over three months and include teachers' responses to the Language Assessment Literacy Survey ($N = 96$), transcripts of five focus group interviews, and transcripts of twelve semi-structured one-on-one interviews.

For quantitative analyses, I computed in JASP v.0.8.3.1 the descriptive statistics for the overall survey, conducted an external review of the language assessment literacy literature, and carried out an exploratory factor analysis (EFA). Subsequently, I compared my results with Kremmel and Harding's to determine the validity of their survey. For qualitative analyses, I used substantive or open coding to discern how/if the participating Uzbekistan EFL teachers are creating relevant and meaningful assessment experiences for their students. The results are discussed in terms of the relationship between the participating Uzbekistan EFL teachers' cognitions, emotions, and activities of language assessment that they report.

INDEX WORDS: Language assessment literacy, Classroom-level assessment, Assessment and testing, Uzbekistan language teacher education, Second language teacher cognition, Assessment motivation

DISCERNING THE LANGUAGE ASSESSMENT LITERACY OF EFL TEACHERS IN
UZBEKISTAN: AN INDIVIDUAL, SOCIAL, AND SOCIOHISTORICAL TEACHER
COGNITION INQUIRY

by

DAVID CHIESA

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2018

Copyright by
David Lawrence Chiesa
2018

DISCERNING THE LANGUAGE ASSESSMENT LITERACY OF EFL TEACHERS IN
UZBEKISTAN: AN INDIVIDUAL, SOCIAL, AND SOCIOHISTORICAL TEACHER
COGNITION INQUIRY

by

DAVID CHIESA

Committee Chairs: Sara Cushing
John Murphy

Committee: Eric Friginal
Kathleen Bailey

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2018

DEDICATION

For Larry:

Silently supporting,

Never expecting,

Only hoping.

Always,

Believing.

Suddenly,

Gone.

This one's for you.

ACKNOWLEDGEMENTS

I want to express my deep gratitude to my dissertation committee members, who have had an immediate impact on this project. I would like to begin with my co-advisors/chairs, Dr. John Murphy and Dr. Sara Cushing, who have shown an incredible amount of patience with portions of the manuscript, and provided constructive feedback that would skillfully move the manuscript through its various phases. I also appreciate their continuous assistance for keeping my progress on schedule by grounding me and facilitating my understanding of the dissertation genre.

I would also like to thank Dr. Eric Friginal for his advice for me to pursue doing empirical research in Uzbekistan. His willingness to give his time so generously to the reading of my comprehensive examinations and the dissertation has been much appreciated.

My academic career in Applied Linguistics would not have started, or even been imaginable, if it was not for my outside dissertation committee member, MA professor/advisor, co-author, colleague, and friend – Dr. Kathleen M. Bailey. During the dissertation process, she provided me numerous opportunities for professional development in Applied Linguistics (e.g., AAAL Program Planning Committee; Ethics Task Force); her continuous support, encouragement, and belief in me has pushed me to be a better scholar.

I would like to extend my deep appreciation to Luke Harding and Benjamin Kremmel for allowing me to use their survey and their preliminary results from their empirical study, *Towards a Comprehensive Empirical Model of Language Assessment Literacy*, for the dissertation.

As this is international dissertation research, I would like to express my deep gratitude to the organizations that approved and funded this study – Uzbekistan and the U.S. Department of State. I would like to extend my thanks to the government of Uzbekistan who gave me

permission to conduct this study, and a deep gratitude to the staff at the Flying High Training site (pseudonym) for the recruitment of teachers and the organization of the focus group interviews.

Special appreciation to my cultural informants, Dr. Ulugbek Azizov, Svetlana Khan, Klara Nazmutdinova, and Nilufar Begigaeva for numerous discussions about the Uzbekistan cultural context, and clarification of some of the transcribed data sources. I would also like to thank people from the U.S. Department of State in Washington D.C. and in Uzbekistan, who have contributed to the success of the implementation of this project: RELO Kelli Odhuu, RELO John Scacco, Ambassador Spratlen, Nodira Meliboeva, Umida Boltaeva, and the entire Public Affairs Section in Tashkent. Additionally, they provided support with the collection of government documents and their translations.

I would like to thank professors from the Department of Applied Linguistics at Georgia State University including Dr. Viviana Cortes, Dr. YouJin Kim, Dr. Ute Roemer, Dr. Diane Belcher, and Dr. Scott Crossley for their assistance in helping me learn more about research methods, theory of language learning, and teaching, which contributed to my knowledge base in Applied Linguistics. I would also like to acknowledge professors from the Middlebury Institute of International Studies at Monterey including, Dr. Lynn Goldstein, Dr. John Hedgcock, Dr. Renee Jourdenais, Dr. Jean Turner, Dr. Peter Shaw, and the late Dr. Leo van Lier who have helped influenced my beliefs about language, language learning and language teaching, and which, have provided me much guidance throughout the PhD, and have influenced my professional identity.

My grateful thanks are directed toward my colleagues Dr. Sarah Goodwin, Dr. Nicole Pettitt, Dr. Stephen Skalicky, Rurik Tywoniw, and Mackenzie Bristow for emotional support, reading over sections of the manuscript, examining data with me, and/or providing me numerous

insights into the results of the project. I would also like to acknowledge my cohort members Dr. James Garner and Dr. Yu Kyoung Shin who, since day one, have been a part of my evolving identity as a scholar. The dissertation would not have been completed if it was not for the support of the Provost Dissertation Fellowship, which provided me opportunity to think and write. My grateful appreciation for the emotional support of Louise Gobron, Sarah Kegley, Dr. Mizuki Mazzotta, Jillian Siehlmann, and my PhD colleagues from all cohorts past and present, and the tireless efforts of the staff in the Department of Applied Linguistics.

Last but not least, I owe a great amount of debt of gratitude for my immediate family members, my mother Elaine and my sister Jennifer, who have been patient through all of my trials and tribulations during the PhD process. Given the recent loss of my father, whom I have dedicated this manuscript to, my mother and sister's emotional support and untiring patience has meant even more. Thank you all.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	V
LIST OF TABLES	XVIII
LIST OF FIGURES	XX
LIST OF ABBREVIATIONS	XXI
<i>Teaching English as a Second Language (TESL)</i>	<i>xxi</i>
<i>Teaching English as a Foreign Language (TEFL)</i>	<i>xxi</i>
<i>Teaching Foreign Language (TFL)</i>	<i>xxii</i>
1 INTRODUCTION.....	1
1.1 Expanding the Domain of Teacher Cognition Studies – Assessment.....	1
1.2 An Under-Researched Population of EFL Language Teachers	2
1.3 Chapter Descriptions	6
2 LITERATURE REVIEW	12
2.1 Teachers’ Decisions and Decision-Making.....	13
<i>2.1.1 Clark and Peterson (1986).....</i>	<i>14</i>
2.2 Conceptual Migration from Education to Applied Linguistics	15
<i>2.2.1 Woods (1996).....</i>	<i>16</i>
2.3 Second Language (L2) Teacher Cognition	17
<i>2.3.1 Emotional Lived Experience</i>	<i>19</i>
<i>2.3.2 Intentionality</i>	<i>21</i>

2.3.3	<i>L2 Teacher Cognition Conclusion</i>	22
2.4	Researching Cognitions of L2 Teachers	22
2.4.1	<i>Individual Ontological Era</i>	23
2.4.1.1	Exemplar Study from SLTC Literature: Karavas-Doukas (1996)	24
2.4.1.2	Methodological issue 1: Survey research.....	25
2.4.2	<i>Social Ontological Era</i>	31
2.4.2.1	Exemplar Study from SLTC Literature: Numrich (1996).....	32
2.4.2.2	Methodological issue 2: Focus groups	33
2.4.2.3	Exemplar Study from SLTC Literature: (Burri, Chen, & Baker, 2017) 35	
2.4.2.4	Methodological issue 3: The semi-structured interview	36
2.4.3	<i>Sociohistorical Ontological Era (2000s onwards)</i>	36
2.4.3.1	Exemplar Study from SLTC Literature: Liu & Fisher (2006)	37
2.4.4	<i>Complex, Chaotic Systems Ontology (2010 onwards)</i>	38
2.4.4.1	Exemplar Study from SLTC Literature: Kiss (2012).....	40
2.5	Concluding Remarks on L2 Teacher Cognition Research	41
2.6	Topics and Contexts Studied	42
2.7	Relationship Among Teaching, Learning, and Assessing	47
2.8	Assessment Literacy	51
2.8.1	<i>Language Assessment Literacy</i>	52
2.8.2	<i>Developing Language Assessment Literacy</i>	54
2.8.3	<i>Identifying the Needs of Stakeholders in Language Assessment</i>	55

2.9	Methodology in LAL Research – Questionnaires and Validation.....	58
2.9.1	<i>Beyond Assessment Needs for Teachers</i>	<i>62</i>
2.10	Methodology in LAL Research – Surveys/Questionnaires + Interviews... 	66
2.11	(Language) Assessment Literacy Conclusion.....	67
2.12	Expanding the Research Domains of LTC and LAL	68
2.13	Conclusion	71
3	SOCIOCULTURAL CONSIDERATIONS – STUDY BACKGROUND.....	73
3.1	Selecting Uzbekistan	73
3.2	History and Culture of Uzbekistan	73
3.2.1	<i>A Historic Moment.....</i>	<i>77</i>
3.3	Establishment of a National In-Service Language Teaching Program.....	79
3.3.1	<i>U.S. Embassy Role</i>	<i>81</i>
3.3.2	<i>Language Teachers at Flying High Training Site</i>	<i>82</i>
3.4	Engagement – Recruitment of participants.....	84
3.5	Conclusion	85
4	METHODS	86
4.1	Research Participants.....	87
4.2	Research Question One	88
4.2.1	<i>Data Collection Material – Adaptation of LAL Survey.....</i>	<i>88</i>
4.2.1.1	<i>Uzbek and Russian Translation of LAL Survey</i>	<i>91</i>

4.2.1.2	Russian Version.....	94
4.2.2	<i>Data Collection Procedure</i>	96
4.2.3	<i>Analyzing the Language Assessment Literacy Survey</i>	97
4.3	Research Questions Two and Three	100
4.3.1	<i>Focus group interviews</i>	100
4.3.1.1	Focus group recruitment and participants.	101
4.3.1.2	Focus group materials	101
4.3.1.3	Focus group procedure	102
4.3.2	<i>Semi-structured interview</i>	103
4.3.2.1	Recruitment and participants.	104
4.3.2.2	Semi-structured interview protocol and procedure.....	104
4.3.2	<i>Analysis of Focus Group and Semi-Structured Interviews</i>	107
4.4	Conclusion	109
5	RESEARCHER POSITONALITY AND REFLEXIVITY	111
5.1	Researcher position in relation to the phenomenon	111
5.2	Researcher position in relation to the training site’s stakeholders	112
5.3	Researcher position in relation to the investigative context and process	114
5.4	Conclusion of Positionality	116
6	RESEARCH QUESTION ONE RESULTS AND DISCUSSION	118
6.1	Overall Numerical Summary	119
6.1.1	<i>Measures of Central Tendency</i>	119

6.1.2	<i>Measures of Dispersion</i>	120
6.1.3	<i>General Categories Numerical Summary</i>	120
6.1.4	<i>Salient Items (Strongly and Not Strongly Endorsed)</i>	121
6.1.4.1	Items that were Strongly Endorsed	122
6.1.4.2	Items that were not Endorsed as Strongly	123
6.1.5	<i>A Visual Representation of Variability</i>	124
6.1.6	<i>How did the participating Uzbekistan EFL teachers respond?</i>	125
6.2	Representation of Taylor (2013)	126
6.2.1	<i>Creating Operational Definitions of the LAL Dimensions</i>	127
6.2.2	<i>Coding Survey Items</i>	128
6.2.2.1	Items of Disagreement Identified by Chiesa and Coders	130
6.2.2.2	Ambiguous Items Identified by Chiesa and Coders	132
6.2.2.3	Comparison: Chiesa and Kremmel/Harding's Initial Coding	135
6.2.3	<i>Summary of Initial Findings</i>	137
6.2.4	<i>Compare: Chiesa's Coding with Kremmel/Harding's Factor Analysis</i> ..	138
6.2.5	<i>Representation of Taylor's (2013) Framework?</i>	139
6.3	Factor Analysis	140
6.3.1	<i>Determining Factorability of Data</i>	142
6.3.2	<i>Choosing a Factor Extraction Method</i>	143
6.3.3	<i>Deciding How Many Factors to Retain and Selecting a Factor Rotation</i>	143
6.3.4	<i>Determining Factor Loadings</i>	145

6.3.5	<i>Interpretation</i>	148
6.3.6	<i>Reporting the Results</i>	149
6.3.7	<i>Chiesa EFA and Kremmel / Harding EFA Comparison</i>	150
6.3.8	<i>EFA Conclusion</i>	155
6.4	Research Question 1 Conclusion	156
6.5	Discussion and Limitations	158
6.5.1	<i>Comparing to Different Stakeholders</i>	158
6.6	Using the LAL Survey for Research and Practice	160
7	RESEARCH QUESTION TWO RESULTS AND DISCUSSION	164
7.1	Assessment-for-Learning (Pre/During Learning)	165
7.1.1	<i>Continuous Assessment</i>	166
7.1.1.1	Continuous Assessment Cognitions	167
7.1.1.2	How Students are Scored on Continuous Assessment	169
7.1.2	<i>Presentations</i>	171
7.1.2.1	Cognitions about Presentations	172
7.1.2.2	How Students are Scored on Presentations	174
7.1.3	<i>Assignments</i>	177
7.1.3.1	Cognitions Around Assignments	177
7.1.3.2	Cognitions about Writing Portfolios	179
7.1.3.3	Cognitions around Peer Assessments.....	180
7.1.3.4	Cognitions around Self Assessments	181

7.1.3.5	How Students are Scored on Assignments.....	182
7.1.4	<i>Participation</i>	184
7.1.4.1	Cognitions about Participation.....	184
7.1.4.2	How Students are Scored on Participation.....	185
7.1.5	<i>Placement and Diagnostic Assessments</i>	187
7.1.5.1	Placement and Diagnostic Assessment Cognitions.....	188
7.1.5.2	How Students are Scored on Placement/Diagnostic Tests.....	189
7.1.6	<i>Conclusion – Assessment-for-Learning (Pre/During Learning)</i>	191
7.2	<i>Assessment-of-Learning (Post Learning)</i>	193
	Note. 53 teachers in total.	194
7.2.1	<i>Final Examinations</i>	194
7.2.1.1	Test Design.....	195
7.2.1.2	Cognitions around Test Design	196
7.2.1.3	Administration and Scoring	203
7.2.1.4	Cognitions around Test Administration and Scoring.....	204
7.2.1.5	How Students are Scored on Final Examinations	205
7.2.2	<i>Conclusion – Assessment-of-Learning (Post Learning)</i>	206
7.3	How do Uzbekistan EFL Teachers Justify Assessment Practices?	208
7.4	Discussion	210
8	RESEARCH QUESTION THREE RESULTS AND DISCUSSION	215
8.1	Sociocultural	215
8.1.1	<i>Collectivism – “Strength is in Unity” (Karimov, 1991)</i>	216

8.1.1.1	Cognitions around Collectivism and Assessment	217
8.1.1.2	Collectivism Shaping Uzbekistan EFL Assessment Responsibilities..	220
8.1.2	<i>Perceptions of Teachers</i>	225
8.1.2.1	Cognitions on Perceptions of Teachers and Assessment	226
8.1.2.2	Perception of the Teacher Shaping Assessment Practices	227
8.1.3	<i>Ethical Responsibility</i>	230
8.1.3.1	Cognitions about Ethics and Assessment.....	231
8.1.3.2	Assessment Responsibilities – Reporting Ethically	234
8.1.3.3	Assessment Responsibilities – Scoring Ethically.....	235
8.1.4	<i>Summary: Sociocultural Considerations</i>	236
8.2	Sociopolitical	238
8.2.1	<i>Presidential Decrees</i>	239
8.2.1.1	Cognitions about Presidential Decrees.....	239
8.2.1.2	Assessment Responsibilities Influenced	241
8.2.2	<i>Outside/Foreign Influence</i>	244
8.2.2.1	The British Council	245
8.2.2.2	Trainings from U.S.-Based Specialists	249
8.2.3	<i>Summary: Sociopolitical Considerations</i>	254
8.3	Overarching Research Question	256
8.3.1	<i>Constraints</i>	257
8.3.2	<i>Affordances</i>	259
8.4	Discussion	262

9	CONCLUSION	267
9.1	Situated EFL Teachers' Cognitions in Uzbekistan.....	269
9.2	The Professional Voice: Identifying Academic Assessment Concepts	270
9.2.1	<i>Implications for Language Assessment Literacy Teacher Education</i>	<i>271</i>
9.3	Generational Voices – Making Sense of Academic Assessment Concepts	274
9.3.1	<i>Implications for Language Assessment Literacy Teacher Education</i>	<i>276</i>
9.4	Cultural Voices: Integrating Academic Concepts	279
9.4.1	<i>Implications for Language Assessment Literacy Teacher Education</i>	<i>280</i>
9.5	Future Directions	283
9.6	Contributions to Academia	284
9.6.1	<i>L2 Teacher Cognition</i>	<i>284</i>
9.6.2	<i>Bridging L2 Teacher Cognition and Language Assessment Literacy</i>	<i>285</i>
9.6.3	<i>Language Assessment Literacy</i>	<i>286</i>
	REFERENCES.....	288
	APPENDICES	309
	Appendix A: Uzbekistan Presidential Decree, No. 24.....	309
	Appendix B: GSU Internal Review Board Approval.....	316
	Appendix C: Adapted Survey in English	317
	Appendix D: Survey in Uzbek and Russian.....	322
	Appendix E: Uzbek Pseudonyms.....	337

Appendix F: Focus Group Protocol.....	338
Appendix G: Semi-Structured Interview Protocol	340
Appendix H: Descriptive Statistics for Each Item	341
Appendix I: Operationalization of LAL Constructs.....	345
Appendix J: Kremmel and Harding, LAL Operationalization of Constructs	348
Appendix K: Researcher’s Finalized Coding Scheme	349
Appendix L: Summarizing Problematic Item Concerns	351
Appendix M: Final Examination Example Sample, PRESETT	352
Appendix N: National Standards – Higher Educational Levels	358
Appendix O: Example Syllabus for LAL for UZ EFL Teachers.....	369

LIST OF TABLES

Table 1 Assessment Literacy Involves..., adapted from Taylor (2009).....	51
Table 2 Levels of LAL, adapted from Bybee (1997), cited from Pill & Harding (2013), p. 383.	54
Table 3 Standards for Teacher Competence (AFT, NCME, & NE, 1990).....	59
Table 4 Research on Language Teachers' Language Assessment Literacy	69
Table 5 Descriptive Statistics of Teachers' Age and Teaching Experience (in years).....	88
Table 6 Items Removed from Original Survey.....	89
Table 7 Uzbek Translation of Survey	92
Table 8 Local Practices Items.....	95
Table 9 Factors based on Kremmel and Harding (forthcoming)	98
Table 10 Focus Group Interview Time Table.....	103
Table 11 Semi-Structured Interview Questions.....	105
Table 12 Semi-Structured Interview Information.....	106
Table 13 Descriptive Statistics.....	119
Table 14 Measures of Central Tendency	119
Table 15 Measures of Dispersion	120
Table 16 Descriptive Statistics – Factors based on Kremmel and Harding (forthcoming)	121
Table 17 Highest Means and Lowest Standard Deviations	122
Table 18 Lowest Means	123
Table 19 Researcher Operationalization.....	128
Table 20 Disagreement Among Coders	130
Table 21 Ambiguity of Items.....	133
Table 22 Chiesa and Kremmel/Harding Coding Comparison I.....	135

Table 23 Chiesa and Kremmel/Harding Coding Comparison II	136
Table 24 Chiesa and Kremmel/Harding Coding Comparison III	137
Table 25 LAL survey: 9 Factor solution (50 items)(Kremmel & Harding, forthcoming).....	138
Table 26 Correlation Matrix	144
Table 27 Items Removed from Factor Analysis to Achieve a More Parsimonious Structure	145
Table 28 Factor Loadings	146
Table 29 Descriptive Statistics – Factors.....	150
Table 30 Assessment-for-Learning (Pre/During Learning).....	166
Table 31 Assessment-of-Learning	194
Table 32 Descriptive Statistics of Teachers' Age and Years' Experience (in years).....	211
Table 33 Research Generations, L2 Teacher Development, and Research Questions	267

LIST OF FIGURES

Figure 1 Test-Operate-Test-Exit [TOTE] decision-making diagram, Shavelson (1973), p. 146 .	13
Figure 2 Teachers' Thought Processes (Clark & Peterson, 1986)	15
Figure 3 Language Teachers' Decision Making (Woods, 1996)	17
Figure 4 Language Teacher Cognition (Borg, 2003a, p. 82)	18
Figure 5 European Social Survey Translation Process	30
Figure 6 Expanded definition of LAL (Fulcher, 2012, p. 126).....	53
Figure 7 Levels of AL/LAL differentiated according to stakeholder constituency	56
Figure 8 Levels of AL/LAL for classroom teachers	57
Figure 9 A conceptual framework of teacher assessment literacy in practice	64
Figure 10 Organizational Structure of Higher Education Institutions	83
Figure 11 Map of Uzbekistan	83
Figure 12 Translation of LAL Survey into Uzbek and Russian	91
Figure 13 Histograms of Item 23 (KT_23) and Item 43 (PBA_43)	125
Figure 14 Researcher Understanding.....	127
Figure 15 Coding LAL Survey	129
Figure 16 Loewen and Gonulal (2015), p. 184. Types of Factor Analyses	141
Figure 17 Scree Plot.....	144
Figure 18 Chiesa, Factor 1: Technical Skills of Assessment w/ Kremmel and Harding's EFA	151
Figure 19 Chiesa, Factor 2 (blue): Social Aspects of Language Assessment; and, Chiesa, Factor 3 (black): Assessment-for-Learning	153
Figure 20 KASA Model.....	277

LIST OF ABBREVIATIONS

The broader umbrella term, L2 teacher cognition, includes the cognitions of language teachers who teach English as a Second Language (ESL), English as a Foreign Language (EFL), and Foreign Languages (FL). Sometimes in the literature, and in daily interactions, people confuse these terms and interchange them often. There are relevant differences among each of these groups of people as language learners and in the types of classes that are taught.

Teaching English as a Second Language (TESL)

TESL classes are commonly offered in countries where English is the predominant language. Students in TESL classes often come from different language and cultural backgrounds, and they may have been in an English-speaking country for different lengths of time. ESL students are diverse. Yet, they usually have ongoing English input in the community, at school, and on the job. ESL teachers can draw upon the real language that students are hearing, reading, seeing, and using in their day-to-day interactions. ESL teachers help students make sense of the language input, so that they can participate with fluency and accuracy in their jobs, at school, and in the community.

Teaching English as a Foreign Language (TEFL)

TEFL classes are most commonly offered in countries where English is not the dominant language. English does not surround the students in school, at work, or on the streets. Students in TEFL classes in public schools usually share a language and an academic course of study. TEFL teachers, textbooks, internet, and online communities can provide the input. In many countries, TEFL teachers are skilled at teaching the language as a subject, and some teachers may still rely on rote learning.

Teaching Foreign Language (TFL)

TFL is a term that is typically used with teachers who teach languages other than English. Often, students in TFL classes usually share a language and an academic course of study. The acronym is often simplified with the language being studied. For example, Chinese as a Foreign Language is referred to as CFL and Japanese as a Foreign Language is identified as JFL. In the United States, the language teachers who teach FL most commonly serve within a modern language department.

1 INTRODUCTION

Second language (L2) teacher cognitions can be complex, ranging over many different subjects; they can be dynamic, changing over time and under different influences; and they can be systematic, forming unified and cohesive personal and practical theories (Feryok, 2010; Johnson & Golombek, 2016). The notion of the mind is thus an important phenomenon to analyze if one wants to understand the process of language teaching. A teachers' mental work extends beyond what can be publically accessible through in-person, audio, or video observation (Burns, Freeman, & Edwards, 2015). For instance, there is a copious amount of private mental work that goes into the planning, evaluating, reacting, and deciding stages of teaching. The mental work remains unobservable to outsiders and beyond the direct reach of researchers (Burns, Freeman, & Edwards, 2015). The language teaching mind is amendable to the fluidity of time – before, during, and after a course of study – and is situated within a specific local setting that is embedded within a larger macro-environmental (e.g., sociocultural and sociopolitical) context. Recently, scholars (e.g., Golombek & Duran, 2014) who want to unravel the complexities of language teachers' mental lives, examine the relationship among cognitions, emotions and activities of the teaching process.

1.1 Expanding the Domain of Teacher Cognition Studies – Assessment

Language Teacher Cognition research is diverse in its subject matter and has been carried out within many different L2 and foreign language (FL) education contexts. Some notable topics are L2 teacher cognitions about grammar teaching and grammatical terminology (Andrews, 1999; Borg, 1999). From this early research, Borg synthesized the field of L2 teacher cognition and operationalized cognitions as what teachers know, think, and believe (Borg, 2003a) about teaching. From his synthesis, topics have expanded to include language teacher identity (Kanno

& Stuart, 2011); language teachers' beliefs and understandings of culture (Feryok & Oranje, 2015); and language as pedagogical content (Breen, Hird, Milton, Oliver, & Thwaite, 2001) to name a few. The subject matters that relate to a language teachers' teaching activities in Language Teacher Cognition research are extensive and will be reviewed in Chapter 2.

For the dissertation, I expand the subject matter addressed within L2 Teacher Cognition tradition by including the subject matter of assessment. The centrality of assessment in teachers' professional practices is argued as not only important but also essential. According to Inbar-Lourie (2013) assessment is always situated within specific institutional and policy contexts and can play a role in a language teacher's instructional practices and professional identity. Language assessment literacy has been reported in the assessment literature to refer to the level of a teacher's engagement with constructing, using, and interpreting a variety of assessment procedures to make decisions about a learner's language ability (Taylor, 2013). This discussion is commonly referred to as the knowledge base of language assessment/testing. Researchers who wish to investigate the cognitions of language teachers related to assessment should expand their research beyond the knowledge base of language assessment/testing to encompass the context, at both the institutional and larger macro-sociocultural levels that the language teacher is a part of. Language assessment literacy for teachers is not a static entity (e.g., having or not having literacy) but the capacity of a teacher to be responsive to the fluidity of one's own cognitions on assessment and testing practices in specific sociocultural contexts in order to get at assessment at the service of learning.

1.2 An Under-Researched Population of EFL Language Teachers

I conducted a study of EFL teacher cognitions on assessment with teachers from Uzbekistan. There are both personal and professional reasons why I carried out the study. Within Central

Asia, I have been particularly interested in its culture, political history, and language planning and policy, all of which have had a great impact on the people of Uzbekistan and the professional lives of EFL teachers. Professionally, as a researcher, I have been cognizant that there are some parts of the world that are studied intensively (e.g., developed countries) and generalized to the rest of the world. Through the study of an under-researched population (i.e., Uzbekistan), I was able to gain a more complete picture of the L2 teacher cognition and language assessment literacy phenomena.

The genesis of the study was that I had an opportunity with the U.S. Department of State to work with EFL teachers in Uzbekistan. I wanted to know how much they knew about assessment so that I could develop an appropriate curriculum for them for the national in-service language teacher education program. The training program was instituted by President Islam Karimov. His Presidential Decree 1875 was entitled, “About measures for further improvement of the system of retraining and advanced training of managerial and pedagogical personnel in higher educational institutions” (See Appendix A for an English version of PD#1875). The decree established that all university teachers and administrators across Uzbekistan must partake in professional development courses once every four years. Since the decree was signed, thirteen cohorts, approximately 1,200 university EFL teachers from Uzbekistan’s twelve provinces, one autonomous region, and an independent city of Uzbekistan have come to the capital, Tashkent, and took part in a TEFL training (see Chapter Three for more information). This training, similar to many L2 teacher education programs, does not explicitly include language assessment as part of the curriculum, as the curriculum focused on knowledge and skills of language and language teaching methodology.

The dissertation begins the process to tap into the mental lives of one cohort of ninety-six individual participating Uzbekistan EFL teachers, by aspiring to understand what they perceive to be valuable in terms of knowledge and skills of assessment, what they do (with assessment), and why they do what they do (Freeman & Johnson, 2005). This research more fully illuminates the conceptual clarity of L2 teachers' language assessment literacy with the analysis of social, cultural, historical, and political factors. Scarino (2013) explains that it is necessary to

consider not only the knowledge base in its most contemporary representation, but also the processes through which this literacy is developed. In line with contemporary, sociocultural learning theories, these processes should recognize the "inner" world of teachers and their personal frameworks of knowledge and understanding and the way these shape their conceptualizations, interpretations, decisions and judgments in assessment. (p. 316)

Language assessment literacy needs to be considered in relation not only to teacher knowledge, but also to teachers' interpretive frameworks, "which are shaped through their particular situated personal experiences, knowledge, understanding and beliefs" (Scarino, 2013, p. 322). To address Scarino's call to action, I investigated Uzbekistan EFL teachers' understanding of the knowledge base of language assessment/testing with the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming). I received permission to use Kremmel and Harding's survey when I met Harding at the Language Assessment Literacy Symposium (2016) in the United Kingdom. He looked for participants to pilot his survey with different populations around the world. In order to support and add to the research being conducted by Kremmel and Harding, I took-up their call for assistance and used their survey in Uzbekistan. Then to tap into the Uzbekistan EFL teachers' personal frameworks of knowledge and address how their specific teaching contexts

shape their decisions and judgments in assessment from an L2 teacher cognition research agenda, I used the interview format (i.e., focus group and semi-structured), which is commonly utilized in teacher cognition research. I present the three overarching research questions (and nine subquestions), which guide the study:

- 1) To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy?
 - A. What assessment skills and knowledge do Uzbekistan EFL teachers believe teachers need to possess?
 - B. Is the theoretical basis for the survey (Taylor's 2013 framework of LAL) faithfully implemented in the survey?
 - C. What is the factor structure present in the Uzbekistan EFL teachers' responses?
- 2) How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students?
 - A. What do Uzbekistan EFL university teachers report they do to assess their students (i.e., assessment tasks and scoring procedures)?
 - B. What are their cognitions surrounding those assessment practices?
 - C. How do their cognitions about assessment shape how they assign scores to their students' work?
- 3) What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students?
 - A. What are the macro-environmental factors Uzbekistan EFL university teachers

report?

B. What are their cognitions surrounding these factors?

C. How do the reported factors and cognitions shape Uzbekistan EFL university teachers' assessment practices?

1.3 Chapter Descriptions

To address the three research questions above, I have laid out the dissertation in the following nine chapters: The Introduction, Literature Review, Sociocultural Considerations to the Study Background, Methods, Research Positionality and Reflexivity, Research Question One Results and Discussion, Research Question Two Results and Discussion, Research Question Three Results and Discussion, and Conclusion. I present a summary of each chapter which is meant to be read in its presented order.

In Chapter 2 – *Literature Review* – I discuss L2 teacher cognition as an area of research inquiry, through an examination of where it came from, where it is now, where it is going, and how this study situates itself within the L2 teacher cognition research tradition. Then I address the topics of assessment literacy (in general) and language assessment literacy (in particular) with a presentation of the different theoretical and practical approaches to their conceptualizations, drawing upon scholarship from the language testing, language assessment, and teacher education communities. Following these orientations, I address research methods, particularly how conceptualizations of and research into L2 teacher cognition have changed over time, as related to ontological stances and methodological approaches, and how language teachers' language assessment literacy has been addressed within studies in language teacher education (regarding assessment). I specifically present the research methods of survey research (questionnaires and validation) and interviews (e.g., focus groups and semi-structured

interviews) because these two methods have been prominent in the literature to analyze L2 teachers' language assessment literacy. The examination of the literature paints a picture of what has been done in the respective fields, and how I propose to move both (L2 teacher cognition and L2 assessment literacy) fields forward.

In Chapter 3 – *Sociocultural Background to the Context of the Study* – I discuss the sociocultural background to the setting of the study, which presents a brief discussion of the social, cultural, historical, and political factors in Uzbekistan. First, I explain in detail why I chose Uzbekistan. Then, I show how Uzbekistan's history, culture, and politics have influenced the development of English language teaching throughout the country. Moreover, I explain Uzbekistan's tenuous relationship with the United States, which has been a key factor to the development and implementation of this study and particularly my role as a researcher (e.g., my positionality). Finally, I explain how the government of Uzbekistan has had an engaged role in this study. This chapter describes the broader conceptual system (Larsen-Freeman & Cameron, 2008) that the participating Uzbekistan EFL teachers' cognitions are situated within.

In Chapter 4 – *Methodology* – I discuss the research methods used to address each research question. These are drawn from the individual, social, and sociohistorical ontologies of L2 teacher cognition research (see Chapter 2). Then, I explain the materials, data collection procedure, and quantitative analysis of research question one and its sub questions. Subsequently, I present research questions two and three, with an explanation of the recruitment process, participants, interview protocols, and qualitative analyses (including intercoder agreement).

In Chapter 5 – *Researcher Positionality and Reflexivity* – I explain my positionality as a researcher in the Uzbekistan context. This chapter is important and is placed before the results

section. The reason for placement there is as follows: I want to be as transparent as possible with respect to my positionality, which had a direct effect on how I collected and interpreted data. My positionality (both emic and etic perspectives) as a researcher, a teacher trainer, a specialist in assessment, and a fellow for the U.S. Department of State all played a role in who I was able to talk to and how the participating Uzbekistan EFL teachers responded to my inquiries as a researcher. The reason for creating a positionality statement cannot be overstated because it informs me, the researcher, and you, the reader, that I am as cognizant as possible of my own biases and I am trying to address them as well as possible as a researcher.

In Chapter 6 – *Research Question One Results and Discussion* – I report the results of the first research question: To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy? To answer this inquiry, I address the following three questions: A. What assessment skills and knowledge do Uzbekistan EFL teachers believe teachers need to possess? B. Is the theoretical basis for the survey (Taylor's 2013 framework of LAL) faithfully implemented in the survey? C. What is the factor structure present in the Uzbekistan EFL teachers' responses? This chapter directly connects the academic fields of L2 teacher cognition and language assessment literacy. I address the individual ontology in L2 teacher cognition research because I use a survey tool created by scholars in the field of language testing, which is meant to discern what the participating Uzbekistan EFL teachers' think is important skills/knowledge of language assessment and testing. Burns, Freeman, and Edwards (2015) explained that the individual ontological research generation in teacher cognition research was essentially "cognitivist, examining the beliefs the language teacher held, how and why these beliefs were constructed, and how they related to practice" (p. 589). I then discuss how the

results match up with theory and research in L2 teacher cognition and language assessment literacy studies. Second, I discuss the limitations of using a survey across cultures. Finally, I discuss survey recommendations to use the Language Assessment Survey with underrepresented population of classroom language teachers around the world.

In Chapter 7 – *Research Question Two Results and Discussion* – I present the findings of the qualitative analyses, which were conducted to address the following research question: How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students? To address this inquiry, I ask the following three sub-questions: A. What do Uzbekistan EFL university teachers report they do to assess their students (i.e., assessment tasks and scoring procedures)? B. What are their cognitions surrounding those assessment practices? C. How do their cognitions about assessment shape how they assign scores to their students' work? These results include a discussion of both formal (i.e., assessments that have points or grades attached to them) and informal assessments (i.e., comprehension checks), but most data sources will exemplify formal assessments. Additionally, these findings will discuss what the participating teachers report that they do in their classes and schools (micro-institutional contexts) to assess their students. Also, I discern what they think, know, believe, and feel about those assessment practices to create meaningful learning assessment situations for their students.

This chapter addresses the social ontology in L2 language teacher cognition research because it focuses on how the wider surroundings, both internal to the person and external to the social setting serve to shape teacher thinking in this area. “The unit of analysis of such studies shifted away from quantification to uncover insights, rather, based on qualitative interpretation and meaning and introduced a move from researcher-determined decisions and beliefs about

language teacher thinking (as opposed to research question one where the participants answered a survey based on what testing specialists believe is important for them to know and do with the language assessment literacy construct) to participant-oriented conceptualizations and explanations” (Burns, Freeman, & Edwards, 2015, p. 591). I then discuss how the focus groups and interviews provided insights into some of the unobservable affective factors that influence the day-to-day assessment practices of the Uzbek language teachers. Also, I discuss other possible avenues to explore the social ontology in L2 teacher cognition research on assessment practices.

In Chapter 8 – *Research Question Three Results* – I present the findings of the qualitative analyses, which were conducted following the research question and subquestions: What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students? To address this inquiry, I asked the following three questions: A. What are the macro-environmental factors Uzbekistan EFL university teachers report? B. What are their cognitions surrounding these factors? C. How do the reported factors and cognitions shape Uzbekistan EFL university teachers’ assessment practices? To answer these questions, I used the same methodology featured in Chapter Seven, in which I interviewed forty-one EFL teachers who participated in a focus group and/or one-on-one semi-structured interviews. Chapter eight addresses the sociohistorical ontology in language teacher cognition research. “Language teaching was shown to occur in situated interactions between teachers’ personal propensities and social practices. Thus, the unit of research analysis in this ontological system placed emphasis on capturing thinking as a function of place and time operating through interaction or negotiation” (Burns, Freeman, & Edwards, 2015, p. 592). Inherent in this orientation was recognition of ways in which my representation of meaning and

their positioning within language teacher cognition research contributes to the process. In the chapter I discuss how the participants referred to each macro-environmental facet as either a constraint or affordance on their assessment practices.

In Chapter 9 – *Conclusion* – I conclude the dissertation and demonstrate how Uzbekistan EFL teachers' cognitions on language assessment literacy are being constructed, negotiated, and conceptualized over time within evolving sociocultural and sociopolitical contexts. First, I locate the EFL teachers' cognitions on language assessment/testing practices within the setting of modern Uzbekistan (i.e., post 1991). Then, I show the relationship between their maturing cognitions and how they identify, make sense of, and integrate assessment concepts into their everyday experiences. Following, I present opportunities and challenges for teacher educators to support the development of EFL teachers' language assessment literacy. Finally, I discuss future directions for research and articulate how the dissertation contributed and extended the academic fields of L2 teacher cognition and language assessment literacy.

2 LITERATURE REVIEW

In this chapter I discuss second language (L2) teacher cognition as an area of research inquiry, through an examination of where it came from, where it is now, and where it is going. Subsequently, I address research methods, particularly how conceptualizations of and research into L2 teacher cognition have changed over time, as related to ontological stances and methodological approaches. I specifically discuss survey research (e.g., questionnaires with validation methods – factor analyses) and interviews (e.g., focus-group and semi-structured interviews) because these methods have been prominent in the literature to analyze L2 teachers' cognitions. Following this timeline, I highlight the diversity of research on language teacher cognition, in terms of the range of topics that scholars have addressed, and then, present the content matter of this study – (language) assessment literacy – with an explication of its different theoretical and practical approaches to its conceptualization. Additionally, I discuss research methods for the academic field of language assessment literacy (e.g., questionnaire design, scale development, and validation). Finally, I show how the study proposes to move the field of L2 teacher cognition forward with (1) an expansion of the subject matter addressed in the Language Teacher Cognition tradition to include assessment; and (2) the use of different epistemological research approaches, which will encompass the sociohistorical paradigm to analyze L2 teacher cognitions of assessment. I also show how the study proposes to move the field of language assessment literacy forward with the (1) support of the validation of Kremmel and Harding's Language Assessment Literacy survey; and (2) contribution to the premise that different stakeholders in language assessment have varying needs.

2.1 Teachers' Decisions and Decision-Making

Education researchers in the 1970s and 1980s identified teacher thinking as decision-making as an important and worthwhile area of inquiry. They defined teacher thinking as the *mental lives* (Walberg, 1972) of teachers. They investigated how teachers' mental lives had been shaped by the activity of teaching in diverse sociocultural contexts (Clark & Peterson, 1986). It is unclear precisely where the idea of teacher thinking as decision-making originated, although related work in general education references Shavelson's (1973) article 'What is the basic teaching skill?' In this seminal discussion, Shavelson alluded to a gap he recognized between the then-prevailing skill-based view of teaching and the ways in which teachers actually work in classrooms. He argued that truly outstanding teachers are not those who can ask the right type of higher-order thinking question, but those who have the *ability* to decide when to ask such questions (Shavelson, 1973). He drew upon then-current cognitive theories of information processing and proposed a model in which a teacher's decisions "flowed logically from making a judgment in a situation, to acting on the judgment, to evaluating if the action had the intended effect" (Shavelson, 1973, p. 151). Figure 1 below is a model intended to depict links between thinking and action.

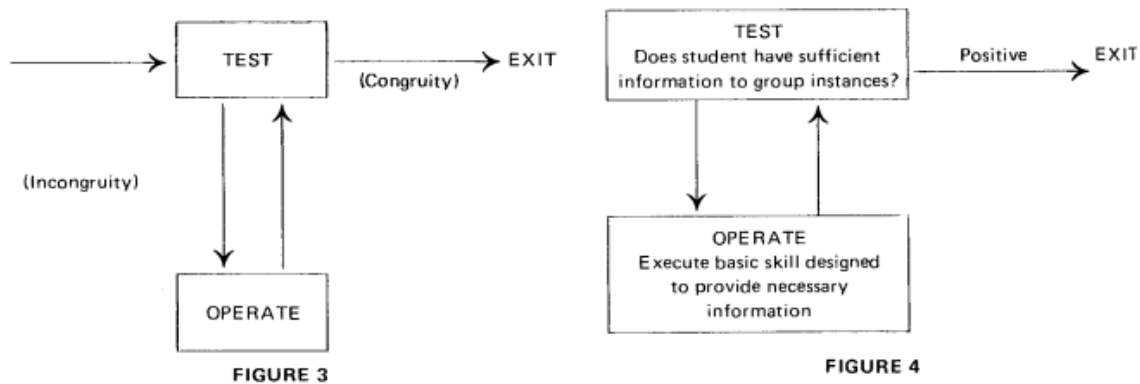


Figure 1 Test-Operate-Test-Exit [TOTE] decision-making diagram, Shavelson (1973), p.

Shavelson's diagram accomplished several important things by effectively locating *where* and *how* thinking happens: in the teacher's judgment of congruity (i.e., the information available from the teacher is equal to or exceeds what a student needs) or incongruity (i.e., the information provided from the teacher is less than what a student needs); or, in other words, teachers' on-the-spot decision-making between intention and action. The thinking process identified by Shavelson (1973) is important for teacher educators, who could now design activities focusing on teachers' decision-making processes. However, as Freeman (2016) explained, a major drawback of the TOTE decision-making diagram is that a teacher's decisions can only be observed when the actions are played out in the real world. In other words, researchers were looking at what could be seen during the process of decision making as opposed to researching thought (non-observable) and action (observable). Thus, within the cognitive, information-processing tradition, teacher thinking was conceptualized as a series of judgments. Not until Clark and Peterson (1986) did the conceptualization of thinking evolve into a more socially-constructed, nested process.

2.1.1 Clark and Peterson (1986)

The connection between teachers' thoughts and the social world was brought more closely together by Clark and Peterson's (1986) model (see Figure 2 below). For these researchers, the process of teaching involves two major areas – first, what is inside the teacher's head (e.g., cognition) and how that interacts with the second area, the social world, along with its observable effects. What I find interesting in Figure 2 is the constraints and opportunities label, because there is a hint of affordance theory (Gibson, 1977), which was not directly mentioned by Clark and Peterson. Affordance theory states that the world is perceived not only in terms of

objects, shapes, and spatial relationships, but also in terms of object possibilities for action (affordances) (Gibson, 1977). In other words, how a person views specific clues in the environment that s/he is a part of (i.e., affordances) will eventually lead to actions that s/he takes. That is, a person's perception drives his or her actions. The concept of affordance thus connects the cognitive with the social in the relationship between the two constructs.

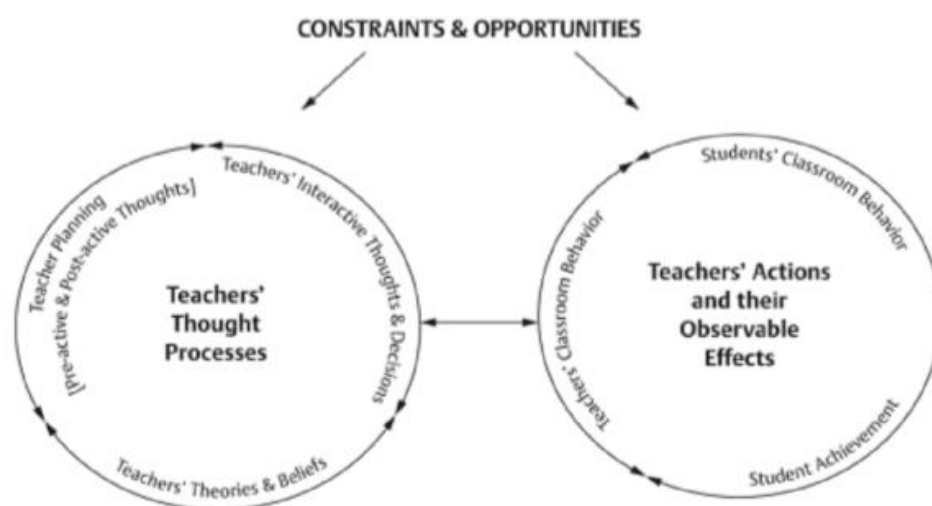


Figure 2 Teachers' Thought Processes (Clark & Peterson, 1986)

Education researchers (e.g., Clark & Peterson, 1986) began to consider questions of teachers' decision-making (perception and action) and other thought processes *combined with and not separated from* clearly observational behaviors such as classroom actions and routines as exceptional pursuits for conducting research, creating/expanding theory, and helping practice.

2.2 Conceptual Migration from Education to Applied Linguistics

From the field of general education, Shavelson (1973) and Clark and Peterson (1986) showed scholars the importance of seeing cognition as a *mental activity*. Woods (1996) and Borg (2003a, 2003b) followed their lead and were influential in the adaptation and development of the

construct of teacher thinking (decision-making processes and general cognition), particularly for *language* teachers. Woods (1996) emphasized the role of the language teacher as a decision-maker, which is similar to Shavelson's position (1973); while Borg (2003a) accentuated the person and role of the language teacher (see Section 2.3), which parallels Clark and Peterson (1986). Now, I discuss what and how Woods (1996) and Borg (2003a), respectively, contributed to the research area of language teacher thinking (as decision-making) and then subsequently language teacher cognitions.

2.2.1 Woods (1996)

Woods conducted a longitudinal study in Canada that focused on a group of teachers and their planning and teaching decision-making processes. In the book's opening pages, Woods remarked "there is still relatively little research on what the second language teacher brings to the process of second language learning... [to date] the role of the teacher has remained a relatively peripheral component" (Woods, 1996, p. 2). The way Woods positioned himself and his thoughts about the role of the language teacher in the process of the student learning shifted the way the field perceived the role of teacher thinking in language teaching, linking the individual teacher to experiences of the students. After drawing on interviews, video-based stimulated recalls, observations, and teacher logs, Woods' (1996) study tracked the process of planning and teaching throughout the teachers' courses. He discovered there are both external and internal factors that contribute to a language teacher's decision making. He created a model (see Figure 3 below) that exemplifies the findings from his study. The model was divided into three distinct and essential time elements. In other words, the dimension of time was an important feature of language teachers' decision-making processes. As summarized by Freeman (2016), "[T]here was a time before the lesson (planning processes), time during the lesson (interactive decisions), and

time after the lesson (interpretive processes)” (p. 156). Woods postulated that the basis for a teacher’s judgments of efficacy were the teacher’s beliefs, assumptions, and knowledge (BAK).

The three different phases of a language teacher’s decision-making process are reflected in the circular model from Woods (1996) as depicted below. This model is similar to Clark and Peterson’s (1986) circles, but differs from Shavelson (1973) in that Woods was trying to create a model to more fully represent the iterative nature of a language teacher’s decision-making.

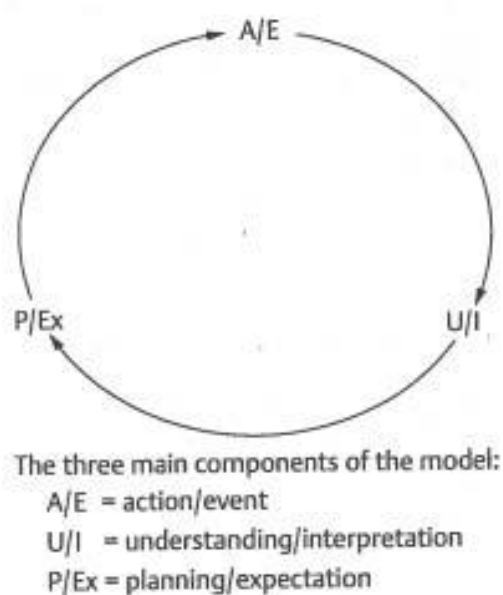


Figure 3 Language Teachers' Decision Making (Woods, 1996)

Woods’ iterative model portrayed decision-making as sense-making within specific settings and activities. Thus, it viewed time and place as important and relevant concepts, which can alter or shape a language teacher’s instructional decisions.

2.3 Second Language (L2) Teacher Cognition

Seven years after Woods' (1996) seminal work, in which he identified teacher thinking as decision making in language teaching from an individualist, cognitivist perspective, Borg

(2003a) took up Woods' work and reconceptualized the concept of teacher thinking as cognitions. Borg's work drew upon the shift within cognitive psychology in understanding learning from a cognitive to a sociocognitive perspective. He explained that teacher cognitions are "the unobservable dimension of language teaching" (p. 81) in which researchers examine cognitions as what language teachers think, know, and believe (Borg, 2006) in the social context. Cognitions are, thus, units of thinking happening at a particular point in time and within a specific context. This understanding underlies the domain of mainstream language teacher cognition (Borg, 2003a.b., 2006, 2009, 2012) research. More specifically, Borg's (2003a) use of the term *cognition* reflects an integration of sources of knowledge, which includes schooling, professional coursework, contextual factors, and classroom practice (see Figure 4).

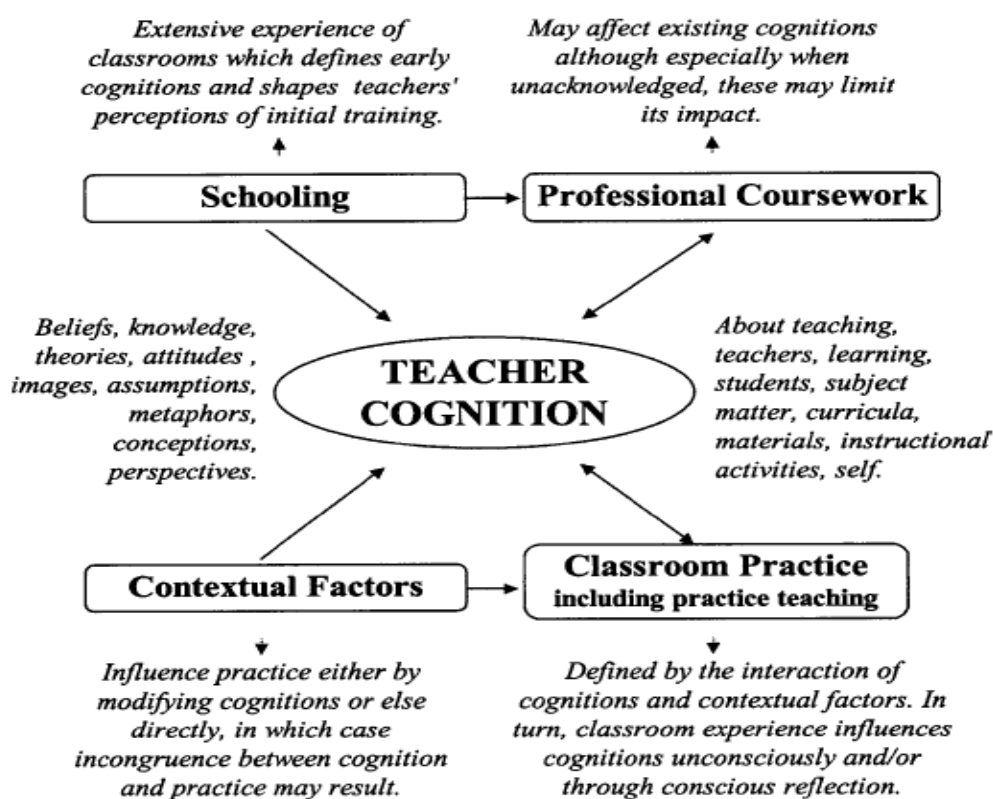


Figure 4 Language Teacher Cognition (Borg, 2003a, p. 82)

The first two categories in the figure, of schooling and professional coursework, are *historical* – concerning past events – because they contribute to teachers’ background and the formation of their thinking (Lortie, 1975). The narrative (or explanatory language) that is created in the descriptions written above and below the four main categories in Figure 4, suggests a specific ordering of time. If one begins at the top left and then moves in a Z-like motion (i.e., from schooling to professional coursework to contextual factors and then to classroom practice), a reader can understand that the figure is designed to connect the past to the present.

Woods' (1996) view of teacher decision-making and Borg's (2003a; 2006) of teacher cognitions played a major role in shaping the operationalization of thinking in language teaching in the 1990s and 2000s. The evolution from decision-making, to thought processes, to BAK (Woods, 1996) and language teacher cognition(s) has shifted how thinking is conceptualized from an individual, internalized process to an interactive, socially grounded one. However, what is not explicitly discussed through this trajectory is the concept of affect and how cognition and affect may be interrelated.

2.3.1 Emotional Lived Experience

Golombek and Doran (2014) and Johnson and Golombek (2016) share many of Borg's (2003a, 2006, 2009, 2012) views on language teacher cognition, but also raise some additional concerns. They explain that Borg offers “a limited epistemological framework for understanding [teacher] cognition, largely informed by the cognitivist paradigm” (Golombek & Doran, 2014, p. 436). Although they frame their discussion within the context of Borg's (2003a; 2006) work, they do not recognize that Borg's later work (e.g., Borg, 2012) acknowledges the recent expansion of L2 teacher cognition to include emotion and identity. Golombek and Doran's position of the language-teaching mind is grounded in Vygotskian Sociocultural Theory (Vygotsky, 1978;

1986). They believe the operations of the human mind are inherently social in nature and that language emerges out of participation in external forms of social interaction (interpsychological) that become internalized psychological tools for thinking (intrapsychological). Although there is no direct link that can be established between the external and internal, from a Vygotskian sociocultural perspective, cognition is mediated through tools and artifacts such as language.

What Golombek and Doran (2014) and Johnson and Golombek (2016) argue is that the dynamic relationship among cognition, activity, and emotion are given insufficient attention in Borg's (2003a; 2006) operationalization of L2 teacher cognition. Those who study language teacher cognition should consider, they believe, the implicit link between cognition and emotion to mediate the interpsychological/intrapsychological (Johnson & Golombek, 2016), because the choice is not whether to feel or not, since emotions are inevitably present in any teaching and learning event. "It is this affective volitional dimension of thought – especially emotions – that Vygotsky (1986) considered as the last 'why' in the analysis of thinking" (Golombek & Doran, 2014, p. 104). Vygotsky acknowledged the relationship between cognition and emotion. Vygotsky (1986) is quoted in DiPardo and Potter (2003), and he, Vygotsky, says that emotions are not a state within a state; instead, "thought is engendered by motivation" (Vygotsky, 1986, p. 252). Vygotsky (1994) addressed the generative capacity of emotions in the development of cognition through the concept of *perezhivanie* – lived or emotional experience. "Learners perceive experiences in a new environment through the prism of *perezhivanie*, a cognitive and emotional reciprocal processing of previous and new experience" (Golombek & Doran, 2014, p. 104).

An emotional subtext is inherent in some of the research on L2 teachers' cognitions (Johnson & Golombek, 2016). From the teacher cognition literature, researchers investigating

novice language teachers have used various constructs that evoke emotional connotations such as *tension* (Freeman, 1993) and *concerns* (Richards, Ho, & Giblin, 1996). Research based in sociocultural theory in language teacher education has addressed teacher emotion overtly through the conceptions of contradictions and emotional dissonance as potential sources of novice teacher learning (DiPardo & Porter, 2003). “Although the construct of teacher cognition appears to validate teachers as thinking professionals, this term semantically maintains distinctions between cognition and emotion... research in language teacher cognition unconsciously reinforces the Cartesian dualism” (Golombek & Doran, 2014, p. 110). Thus, L2 teacher cognition research should not only address Borg’s (2003a; 2006) operationalization, but also needs to include the emotional lived experience of the teacher, which would unite interpsychological and intrapsychological considerations.

2.3.2 *Intentionality*

Moving forward in the teacher cognition research agenda, Kubanyiova and Feryok (2015) expressed a desire to redraw the conceptual landscape of language teacher cognition in applied linguistics research, because a social alternative to the cognitivist epistemological perspective as stressed by Borg (2003a; 2006) seems better suited to address their larger vision:

to embrace the complexity of teachers’ inner lives in the context of their activity and aspire to understand what we have broadly termed ecologies of language teachers’ inner lives, as these relate to what language teachers do, why they do it, and how this may *impact how their students learn* [emphasis mine] (Kubanyiova & Feryok, 2015, p. 436).

The larger vision situates the end goal in mind – understanding how language teachers create learning experiences for their students (Freeman & Johnson, 2005). Kubanyiova and Feryok (2015) emphasize that L2 teacher cognition research should focus on the process of *intentionality*

– a focus on purposeful human actions, which are both individual and collective (Searle, 1990) and include an amalgamation of the three mental faculties of cognition, emotion, and motivation (Schweikard & Schmid, 2013). In other words, “Intentionality offers a core concept that links individuals and others, minds and actions, and encompasses the link between teaching and learning” (Kubanyiova and Ferok, 2015, p. 441). Intentionality, thus, is a construct used by L2 teacher cognition scholars to research the inner lives of teachers in action to establish connections to student learning.

2.3.3 L2 Teacher Cognition Conclusion

L2 teacher cognition as an area of inquiry examines what language teachers know, think, believe, (Borg 2003a; 2006), and feel (Golombek & Doran, 2014; Johnson and Golombek, 2016). L2 teacher cognition is conceptualized as an interactive, socially grounded process, the aim of which is intentionality (Kubanyiova & Feryok, 2015). The dissertation study focuses on expanding the domain of language teacher cognition research and will begin the process to try to unravel the complexity of teachers’ inner lives by aspiring to understand what language teachers perceive to be valuable knowledge and skills of assessment, what they do with assessment, and why they do what they do (Freeman & Johnson, 2005). Specifically, I will explore how language teachers’ knowledge, beliefs and feelings about assessment and testing serve to get at assessment for the service of learning. I will utilize research methods from L2 teacher cognition studies to address the phenomenon of L2 teacher cognition and the content matter of assessment.

2.4 Researching Cognitions of L2 Teachers

Research into L2 teacher cognition has changed over time, especially as related to ontological stances and methodological approaches (Burns, Freeman, & Edwards, 2015). Burns, Freeman, and Edwards (2015) identified four different ontological research traditions: (1) the individualist

[1990 onwards] – a cognitivist ontological stance grounded in teachers’ decisions and decision-making practices, thoughts, and beliefs, which tended to be analyzed through quantitative means; (2) the social [1995 onwards] – a sociocognitive paradigm in which researchers analyze the wider surroundings and how the context shapes or informs thinking, analyzed mostly through qualitative means such as using diary studies; (3) the sociohistorical [2000 onwards] – “thinking as a function of place and time, through interaction and negotiation with social and historical contexts” (Burns, Freeman, & Edwards, 2015, p. 589), which has often been researched through qualitative measures, including interviews and narrative inquiry; and finally, (4) the complex/chaotic systems [2010 onwards] – a dynamic and emergent system that involves the integration of multiple interconnected elements, and has been researched (although the number of studies is few) qualitatively through analysis of interactions. These four ontological research paradigms present an overarching conceptual scheme through which the various periods, methods of data collection, and resulting methodological themes of existing language teacher cognition research may be perceived and discussed. However, the four research traditions are not as distinct and discernable as Burns, Freeman, and Richards (2015) describe them. The meta-organizational view of teacher cognition research is meant to provide a road map of the main conceptual features over time.

2.4.1 Individual Ontological Era

The individualist ontology in language teacher cognition research is grounded in teachers’ decisions, thoughts, and beliefs and has been taken up mostly in a quantitative research paradigm. This period is essentially cognitivist (e.g., Farrell, 2009; Richards, 1996; Richards & Nunan, 1990). The analytical unit in this phase of research is the decisions and decision-making processes that could be discerned from a teacher’s teaching practice. The methodology used is

dominated by the quantitative paradigm, but there are also traces of qualitative research and mixed-methods methodology. This period consists primarily of *questionnaires/surveys* – elicitation devices that “obtain a snapshot of conditions, attitudes, and/or events of an entire population at a single point in time by collecting data from a sample drawn from that population” (Nunan & Bailey, 2009, p. 125) (e.g., Baker, 2014; Borg, 2003a; Mattheoudakis, 2007; Urmston, 2003); *attitude scales* (Karavas-Doukas, 1996); *observations* – the act or practice of noting and recording facts and events (Baker, 2014) (in this ontological era, observation research is mostly quantitative); and *stimulated recall interviews*, a procedure in which it “is assumed that the use of some tangible (perhaps visual or aural) reminder of an event will stimulate recall of the mental processes in operation during the event itself” (Gass & Mackey, 2000, p. 17) (e.g., Baker, 2014; Mullock, 2006). The focus of this research tradition is on the teacher and his/her decisions, thoughts, and beliefs. Most teacher cognition studies (see section 2.6 below) that examine L2 teacher cognitions and the teaching of grammar and grammatical knowledge have focused on this individual ontological research tradition, but there are some that have measured different content areas, such as L2 writing.

2.4.1.1 Exemplar Study from SLTC Literature: Karavas-Doukas (1996)

Karavas-Doukas (1996) is an exemplar study that showcases the individual ontological tradition, which focuses on teachers’ decisions, thoughts, beliefs, and attitudes. She utilized a mixed-methods research design to address these teachers’ cognitions. Karavas-Doukas acknowledged that language teachers do not easily adopt a new method/approach in language teaching because they often have difficulty changing their attitudes during a pedagogical innovation. She decided to understand language teachers’ attitudes toward the communicative approach in the 1990s within the context of the Greek public school system. This educational system adopted a series of

textbooks and a curriculum advocated by the Council of Europe Project No. 12 that focused on enhancing and developing students' linguistic and sociolinguistic skills, including their interactive strategies as well as the promotion of students' intellectual and social development. The researcher first developed a Likert-type attitude scale with 85 statements (40 favorable and 45 unfavorable) about aspects of communicative approaches to L2 instruction, and then conducted reliability and validity measurements. Through statistical analysis and editing, Karavas-Doukas distributed the final attitude scale of 24 statements to 101 Greek secondary school English language teachers, 14 of whom were observed in their classrooms and interviewed. The participants' results were analyzed using descriptive statistics and compared to her results from classroom observations and interviews. She discovered that teacher's classroom practices (with her observations) do not necessarily match their expressed favorable/positive attitudes toward communicative approaches, as identified from the attitude scale. Karavas-Doukas (1996) identified that the attitude scale was a cost-effective and easy-to-administer instrument for gathering data on teachers' beliefs on the communicative approach. Although the use of an attitude scale was novel in order to explore the insights into the exact nature of the teachers' attitudes, further clarification and support was needed, especially when the teachers' responses were compared to what they did in their classrooms and said in their interviews.

2.4.1.2 Methodological issue 1: Survey research.

Surveys and questionnaires are used in this conceptual era and are considered *elicitation devices* – tools that cause people to do or say something, eliciting or evoking information from the study participants (Nunan & Bailey, 2009). Dörnyei (2003) emphasizes that surveys are well suited for asking factual questions, behavioral questions, and attitudinal questions. There is an extensive range of items that can be asked in a survey.

Brown (2001) operationalized *questionnaires* as “any written instruments that present respondents with a series of questions or statements to which they are to react, either by writing out their answers or selecting from among existing answers” (p. 6). In daily conversation, people often use the terms *survey* and *questionnaire* interchangeably. Broadly speaking, *survey* is the macro-term, which refers to data collection that can be conducted in writing through a questionnaire or orally through interviews. The types of interviews can consist of focus-group interviews, face-to-face interviews, telephone calls, computer-assisted telephone interviewing, and interactive voice-response. The strengths and weaknesses of each approach, including how to conduct the interviews, have been discussed extensively in the research methods literature (Denzin & Lincoln, 2005; Wolf, Joy, Smith, & Fu, 2016).

In survey research, the researcher does not intervene with objects or subjects in research, aside from observing or asking them to provide data. For instance, Urmston (2003) reported on the results of a survey that was conducted twice – the first was in 1994 and the second was 1997. The survey was given to 40 Bachelor of Arts (BA) students (30 completed the survey) who were studying TESOL at Hong Kong City University. The purpose of the survey was to assess the pre-service teachers’ attitudes toward the areas of language use, lesson planning and decision-making, teaching approaches, professional relationships and responsibilities, and perceptions and values in English language teaching. The survey was analyzed quantitatively between time one and two and found that, as the teachers progressed through the BA courses, they came to realize that there was a conflict between what they were learning in the program and their previously-held beliefs/attitudes toward key issues/concepts in English language education. This survey research has been particularly noted in the teacher cognition literature because it was conducted at two different points in time.

2.4.1.2.1 Survey Development

The challenge for survey researchers is to design questionnaires that capture the information the researchers wish to elicit without impacting the quality or content of that information.

Questionnaires are perceived as easy to construct (Reid, 1990; Wolf, Joy, Smith, & Fu, 2016); however, after careful consideration of the values that underlie them, the amount of time necessary for item construction and validation, and the use of appropriate analyses to answer the proposed research questions, one can see the inherent complexity of trying to design items that address the specific construct(s) being measured (e.g., attitudes, beliefs, opinions, and/or knowledge about something). Research manuals in applied linguistics repeatedly emphasize the importance of validating questionnaires (e.g., Hatch & Lazaraton, 1991), but only a handful of articles in the L2 teacher cognition literature discuss or report on validation results (e.g., Kim, 2015). Through validation, Petric and Czarl (2003) argue that “such an instrument would enable researchers to compare findings in different contexts” (p. 188).

Multiple research method manuals (e.g., Brown, 2001; Cumming & Berwick, 1996; Hatch & Lazaraton, 1991; Seliger & Shohamy, 1989) describe different aspects of validity: content, construct, and response validity. *Content validity* refers to a review of the survey items by subject-matter experts and verification that its content represents a satisfactory sampling of the domain. Usually irrelevant items are discarded, related statements are collapsed, and potential wording of problems are addressed. The second validation process is called establishing *construct validity*, which is “a mutual verification of the measuring instrument and the theory of the construct it is meant to measure” (Cumming & Berwick, 1996, p. 26). Procedures for establishing construct validity include a statistical procedure such as a *factor analysis* – the grouping of variables into clusters according to common underlying factors, which shows

whether the construct behind the instrument is homogeneous or multidimensional (Petric & Czarl, 2003). The final validation process is identified as *response validity* – where participants are asked to explain their reactions and answers to the survey items. The aims of this validation process are to check the wording and interpretation of the items in order to get feedback on the salience of parts and sections of the survey, and to check the survey scale's numbers or descriptors.

Of L2 teacher cognition studies that use surveys to obtain data, many report reliability measurements, but very few report on validity. Kim (2015) developed and validated a survey to measure early childhood (EC) language teacher knowledge, which drew from Grossman's (1990) 'model of teacher knowledge' and included subject-matter knowledge, general pedagogical knowledge, pedagogical content knowledge, and knowledge of context. Kim used descriptive statistics, Cronbach's alpha, and split-half reliability to analyze the reliability and validity of her newly-formed survey. Reid (1990) provides an early illustration of the complexity of trying to design a questionnaire in her seminal article, *The dirty laundry of ESL survey research*. Reid (1990) explained that she "did not realize at the time ... how long and complex the process of norming a new survey instrument could be" (p. 324). Reid's (1990) article set the stage for survey researchers in TESOL to follow specific guidelines for publishing results, which includes reliability and validity measurements (Mahboob et al., 2016).

2.4.1.2.2 Using surveys across cultures – international comparative research

The complexity of survey development increases when data collection instruments are to be used across cultures. *International comparative research* employs strategies that include comparisons across national, societal, and cultural boundaries conducted within international settings, most often by international teams, and can be referred to as 3MC (multicultural, multinational and

multilingual) studies (Hantrais, 2009). The goal of 3MC surveys is to produce comparable measures across multinational, multicultural, or multiregional populations. The paramount challenge is to determine the optimal balance between the local implementation of a design within each country or culture that will also optimize comparison across countries or cultures (Pennell, Cibelli Hibben, Lyberg, Mohler, & Worku, 2017). One of the largest challenges in the design of cross-cultural survey studies is the comprehension of items and composing proper translations that can produce comparable data across cultures (Behr & Shishido, 2016).

2.4.1.2.2.1 Expected Translation Process

According to the Survey Research Center (2017), the survey researcher needs to organize two or more different teams to conduct team translation (see Figure 5 below), in which there are multiple stages. Figure 5 is based on the European Social Survey Translation guidelines (since 2010) and is read from top to bottom. Overall, there are two groups of translation teams that can consist of multiple people per team, with different strengths (Beaton, Bombardier, Guillemin, & Ferraz, 2002; Mohler, Dorer, de Jong, & Hu, 2016). Each team produces a translated version from the source survey and then both teams will come together for the review process, the adjudication, translated verification, Survey Quality Prediction (*SQP coding*) – coding that prevents deviations between the source questionnaire and the translated versions by checking the formal characteristics of the items (Saris, van der Veld, & Gallhofer, 2004), pretesting, and the target instrument.

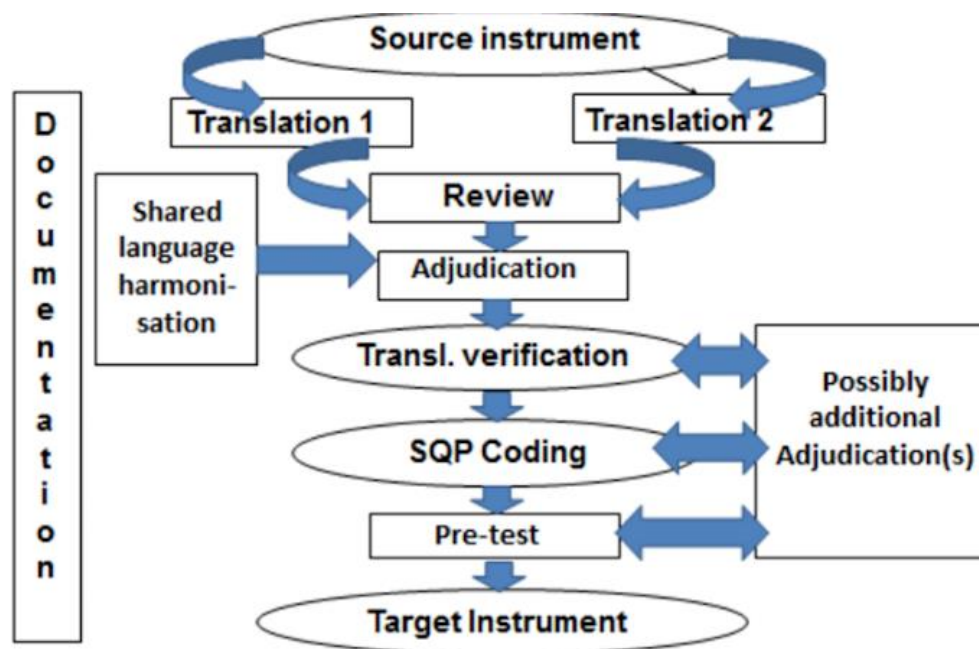


Figure 5 European Social Survey Translation Process

As we move through the figure from the top to the bottom, we can note that each stage of the process builds upon the previous ones and uses the documentation to inform the following stage. Team approaches to translation have been proven to provide “the richest output in term of (a) options to choose from for translation and (b) a balanced critique of versions” (Acqandro, Jambon, Ellis, & Marquis, 1996, p. 575). Many 3MC studies have shown that back-translations, although a popular technique in the social sciences, provide only limited, potentially misleading insight into the quality of the target language text (Harkness, 2008). These scholars recommend that, instead of consulting two source language texts, it is much better in practical and theoretical terms to focus attention on first producing the best possible translation and then directly evaluating the translation produced in the target language, rather than indirectly through a back-translation.

All in all, 3MC studies should utilize the team translation process to overcome one of the largest challenges to the design of cross-cultural survey studies, which is the comprehension of items and composing proper translations that can produce comparable data across cultures.

2.4.2 Social Ontological Era

The social ontological era's conceptual unit of analysis expands from individual teachers' epistemological views into *meanings that are situated in social contexts* (e.g., Tsui, 2003). Its related research methodology is framed not in quantitative but mostly qualitative research methods, including *introspective methods* – “the process of observing and reporting one's own thoughts, feelings, reasoning processes, and mental states” (Nunan & Bailey, 2009, p. 285), which features a range of research methods including stimulated recalls (e.g., Yuan & Lee, 2014); *diary studies* – first-person accounts of a teaching experience that are documented through regular, candid entries in a personal journal (e.g., Numrich, 1996); and, *interviews (focus groups, one-on-one)* – interviews with participants that can last for extended periods or multiple times (e.g., Feryok, 2008; Kubanyiova, 2012; Warford & Reeves, 2003). With the social ontological tradition, research on language teaching moved from identifying *what* teachers think, know, and believe (Borg, 2003a) to understanding *how* shifts in cognition happen through the process of learning to teach across professional careers, and within instructional contexts *where* those learning processes unfold (Freeman & Richards, 1996). Here, a combination of emic (insider) and etic (outsider) perspectives is used. The second era is therefore characterized as social because researchers are looking at *the conceptual changes* in thinking from a sociocognitive perspective, “by emphasizing how the wider surroundings or contexts, both internal to the person and external to the social setting, shapes and/or informs thinking” (Burns, Freeman, & Edwards, 2015, p. 591). In sum, teacher cognition researchers gained a conceptual

shift with this era through which L2 teacher learning was viewed as socially contingent on knowledge of self, students, subject matter, curricula, and settings. An exemplar study from the teacher cognition literature is Numrich (1996).

2.4.2.1 Exemplar Study from SLTC Literature: Numrich (1996)

Numrich (1996) is an example of a study underpinned by a social rather than an individualist ontology. To explore how pre-service language teachers learn how to teach, Numrich asked twenty-six American student teachers who were enrolled in her MA TESOL practicum course to teach four hours per week outside of the university with a co-teacher. For every hour of teaching they were required to generate a reflective journal in which they wrote about their personal experiences and feelings about teaching. Additionally, the student teachers were required to compose a language learner (L2) autobiography. The data generated through their reflective journals and L2 autobiographies were then examined qualitatively to identify recurrent themes. Numrich identified four themes. First, Numrich found the student teachers to be preoccupied with their own teaching during the initial stages of their practicum teaching. Their preoccupations included concerns with making the classroom a safe and comfortable environment and the need to be creative and varied in teaching. Second, the data illustrated that the student teachers were making connections (consciously and unconsciously) with what they said in their language learning (L2) histories and their journals. Third, they had unexpected discoveries about effective L2 teaching, such as an awareness that positive learning can also take place outside the classroom, and that students want higher levels of error correction on pronunciation and grammar than the student teachers had anticipated. Finally, the student teachers experienced continuing frustrations with some of the more mundane classroom management dimensions of the teaching processes such as managing class time, giving clear

directions, and assessing students' learning progress. Numrich posited that such insights were better facilitated through benefits associated with the inclusion of reflective journaling as a research tool for data collection because written journals make it possible for a researcher to gain "insights into some of the unobservable affective factors" (p. 148) of the learning-to-teach process. Numrich's (1996) study exemplifies how social and affective factors contribute to teacher cognition.

2.4.2.2 Methodological issue 2: Focus groups

A *focus group interview* (FGI) is an interview with a relatively small group of people (e.g., a subset of the participating teacher population) on a specific topic. The size of the focus group can range in number and in length of time. A common size for a focus group is six to ten people, and FGIs typically last from thirty minutes to two hours (Patton, 2002). "The twist is that, unlike a series of one-on-one interviews, in a FGI participants get to hear each other's responses and [have opportunities] to make additional comments beyond their original responses as they hear what other people have to say. However, participants need not agree with each other or reach any kind of consensus... the object is to get high-quality data in a social context where people can consider their own views in the context of the views of others" (Patton, 2002, p. 386). Thus, a researcher can obtain data (e.g., perceptions, attitudes, beliefs, reasons, etc.) simultaneously from several interviewees with similar backgrounds (see exemplar teacher cognition study below: Burri, Chen, & Baker, 2017).

Like all forms of data collection, there are both advantages and limitations to using focus groups. Kreuger (1994) explains four advantages to the inclusion of FGIs as a component of qualitative inquiry: (1) Data collection is cost-effective with a much larger sample size in a shorter amount of time than individual interviews; (2) interactions among the FGI participants

enhance the quality of individual and group responses; (3) the extent to which there is a relatively consistent, shared view or great diversity of views can be quickly assessed; and (4) participants tend to enjoy focus groups since the experience taps into human propensities for social interaction. Sociologist Madriz (2000) also iterates that focus groups allow access to research participants who may find one-on-one, face-to-face interactions scary or intimidating. Therefore, when handled well, FGIs can help participants enter into a safe environment where they are better able to share ideas, beliefs, and attitudes.

There are a few limitations to the use of FGIs as a tool for data collection. First, the number of questions that can be asked tends to be restricted in a focus group setting. Kaplowitz (2000) posited that a general rule of thumb is to plan to ask no more than ten major questions. Second, people who might believe that their viewpoints reflect a minority perspective may be reticent to speak up for fear of risking negative reactions from the other FGI participants. Thus, the role of the facilitator is important to ensure everyone's opinion is heard. I use the term *facilitator* instead of *interviewer* because *interviewer* implies a role that is more limited scope (e.g., unidirectional) while the term *facilitator* implies a role that is multidirectional among the person leading the focus group and all of the FGI participants. "The focus group is not a collection of simultaneous individual interviews, but rather a group discussion where the conversation flows because of the nurturing of the moderator" (Kreuger, 1994, p. 100). Third, confidentiality cannot be assured in focus groups even if all of the participants sign an agreement that they will keep personal information confidential. Participants who partake in FGIs have personal agency outside of the FGI, and the moderator cannot control if they divulge information that was discussed. Finally, the discussion leadership capability of the facilitator in conducting a focus group interview requires considerable skills of managing a group beyond simply asking questions. The moderator

must manage the interview so that one or two people do not dominate it, and so that those participants who tend not to be highly verbal are able to share their view. (I will discuss briefly my experiences using FGIs in Chapter 4.)

2.4.2.3 Exemplar Study from SLTC Literature: (Burri, Chen, & Baker, 2017)

A teacher's identity has been argued to convey the "other dimensions of inner lives (e.g., emotions, motivations, values)" (Kubaniyova & Feryok, 2015, p. 436). A recent article by Burri, Chen, and Baker (2017) brought together teacher identity and teacher cognition to theorize student teachers learning, focusing specifically on learning to teach English pronunciation. The fifteen post-graduate students who participated in the study took part in a graduate-level seminar on teaching pronunciation and prosody. Multiple data sources were collected to examine the relationship between cognition development and identity construction in learning to teach pronunciation: a pre-/post-course questionnaire, FGIs, observations of the students in the post-graduate pronunciation class, and semi-structured interviews. The FGIs were important in that they elicited participants' cognitions about pronunciation instruction and critical incidents from the coursework. The critical incidents were discussed to elicit the participants' perceptions and viewpoints on any memorable, unexpected, or challenging experiences during the course. The participants were grouped according to their language background and teaching experience, and each focus group contained of three to five members. Burri, Chen, and Baker (2017) identified themes throughout the data sources, and discovered that identity construction – "manifested through imagination of self and others, engagement and investment in the course, and alignment with the course content – not only had a profound impact on participants' cognition development, but that these two constructs were intertwined in a complex and reciprocal relationship, fostering the process of student teachers' learning to teach pronunciation" (Burri,

Chen, & Baker, 2017, p. 128). Burri, Chen, and Baker's (2017) article is an exemplary study that could be conceptualized as the social ontological research generation which used FGIs (and semi-structured interviews) that elicited participants' cognitions about pronunciation instruction.

2.4.2.4 Methodological issue 3: The semi-structured interview

The semi-structured interview has been used by L2 teacher cognition researchers (as seen above in Burri, Chen, & Baker, 2017). During semi-structured interviews, an interviewer has a general idea of the specific themes to be explored, along with a list of questions that will serve to frame each interview. The lists of questions are the same for each interview; however, the interviewer may also depart from the set of pre-determined questions depending upon what the interviewer perceives to be the relevance of themes as they are emerging, and will not necessarily be tightly constrained by the list of preset questions. Dowsett (1986) posits such flexibility to be a particular strength of semi-structured interviews as a data collection procedure. The format of a semi-structured interview offers the researcher the flexibility of an open or more loosely-structured conversation while also providing a clear focus for the topics being discussed. Nunan and Bailey (2009) explain that the main difference between a semi-structured interview and an unstructured interview "is that the former will adhere more closely to the researcher's agenda than the latter" (p. 314). In essence, as the interview unfolds, the specific topics discussed will serve to assist the interviewer in settling on possible directions for the interview to take, and thus uncover rich descriptive data on the personal experiences of participants.

2.4.3 Sociohistorical Ontological Era (2000s onwards)

A sociohistorical perspective follows the social in the early 2000s and views the research into the minds of language teachers as "thinking as a function of place and time, through interaction and negotiation with social and historical contexts" (Burns, Freeman, & Edwards, 2015, p. 589). This

perspective is multifaceted and multilayered and includes time as a major variable in how research is conducted. As we move forward from individual orientations to social orientations and now to a sociohistorical orientation, we can see a more inclusive view of the language teacher mind, one with ties to Borg's (2006) view of L2 teacher cognition as an integration of sources of knowledge (e.g., schooling, professional coursework, contextual factors, classroom practice; see Figure 4). A Vygotskian sociocultural framework of mind is adopted in this ontological paradigm (Johnson, 2009), which emphasizes how language teaching occurs in situated social interactions between teachers' personal propensities and social practices. A qualitative research approach is mostly used with and through a co-constructed researcher-participant dialogue (e.g., Breen et al., 2001). Other studies in this paradigm include Cross (2010), Liu and Fisher (2006), Hiver (2013), and Golombek and Doran (2014).

2.4.3.1 Exemplar Study from SLTC Literature: Liu & Fisher (2006)

Liu and Fisher's (2006) multiple case study investigated change in four modern foreign language student-teachers' conceptions of self in their classroom performances, teacher identities, relationships with pupils, and self-images in pupils' eyes. Liu and Fisher not only explored the changes in the areas listed above, but also looked at how the student teachers explain the change or lack of change in their conception of self. The data collection process lasted nine months and extended over three different terms. The data included reflective teaching logs, open-ended questionnaires, written reflections, and semi-structured interviews. Intentional variability in data collection methods at different points in time of the research period was considered essential for capturing change at different stages of the process. These data sources were analyzed using *hermeneutic interpretive analysis* – the interpretation of meaning that included the participation of the understander/interpreter in the construction of meaning (Patton, 2002). Change in

conception of self, where classroom performance is concerned, was common to all four teachers and was consistent throughout the course. Also, student-teachers' identities showed consistent positive change with such comments as *real teachers* and *genuine teachers*. The student-teachers' concept of self in relationships with pupils was also straightforward for two of the participating teachers. One teacher "experienced change in her relationships with pupils across the year with a dip occurring on entering a new school context, to which she had to adapt her expectations to of pupil behavior. Self-image in pupil's eyes varied" (p. 358). Findings did not draw direct causal explanations. One noteworthy aspect for Liu and Fisher was that the student teachers hardly ever commented on the core subject of their class – the foreign language. Overall, this study illuminated developmental patterns in a longitudinal study of teachers' conceptions of self; and thus, addressed time (e.g., history) as a major component of the development of the L2 teacher's mind. As such, Liu and Fisher posit the construct of time as a potentially important area for future teacher cognition research.

2.4.4 Complex, Chaotic Systems Ontology (2010 onwards)

Larsen-Freeman's (1997) seminal article into processes of second language acquisition addressed the need for applied linguists to use a complexity science framework for studying how one acquires an additional language. She pointed to Gleick (1987) who stated that "to some physicists chaos is a science of process rather than state, of becoming rather than being" (p. 5). This dynamic and complex system is believed by Larsen-Freeman (1997) to be meaningful for languages and therefore for the study of second language acquisition. The purpose of the article was to show similarities and differences among complex linear systems in nature and language acquisition, not by empirical evidence *per se*, but through analogies and metaphors. In essence, Larsen-Freeman (1997) hoped "that learning about the dynamics of complex nonlinear systems

will discourage reductionist explanations in matters of concern to second language acquisition” (p. 142), which occurs far too often in the social sciences.

The language teaching profession, heavily influenced by areas within applied linguistics such as second language acquisition, took up the construct of dynamic, complex, and adaptive systems for the teaching and teacher education professions. There are six main characteristics of complex adaptive systems that are appealing to language teacher cognition researchers. First, complex systems are heterogeneous in that they are composed of different components or agents. Second, complex systems are dynamic and evolve over time (de Bot, 2008). Third, the complex systems are sensitive to initial conditions, and when there is a change in the system, the change could produce unpredictable, non-linear outcomes. Fourth, complex systems are contextualized, in which they are related to other systems. Fifth, the complex systems are co-adaptive/open, which means they are responsive to and influence change in other systems. Six, complex systems are self-organizing despite the appearance of chaotic patterns (Ellis & Larsen-Freeman, 2006). These six factors of dynamic, complex and adaptive systems could provide language teacher cognition researchers with new lenses to examine the relationships between beliefs, knowledge, and practice.

Extending some of the complexity themes Larsen-Freeman first introduced, Feryok (2010) took up the challenge of connecting the theory of complex, dynamic, adaptive systems to language teacher cognition research. She conducted an exploratory investigation of an already-published article using complexity science with the case study of the practical theory of an English language teacher teaching EFL in Armenia. Her article presented evidence of heterogeneity, dynamics, sensitivity to initial conditions, contextualization, openness, and adaptation. Feryok (2010) concluded that “complex system theory is compatible with other lines

of research, is able to be developed in field-specific ways, offers several lines of research as well as different methodological approaches, and has practical implications for language teacher development” (p. 272).

From 2010 onwards, this ontological generation was defined as complex/chaotic, and utilized a qualitative research methodology including analysis of social, cultural, historical, and political factors (Burns & Knox, 2011; Feryok, 2010; Finch, 2010; Kiss, 2012). This approach to teacher cognition research seems quite different from other ontological approaches in that it tries to establish conceptual clarity among all factors that contribute to teachers’ cognitions. To date, based on what I believe to be a comprehensive review of relevant literature, there are six articles that adopt a complex, chaotic systems ontology for studying the language teaching mind. There is one conceptual article (Feryok, 2010) and five empirical studies (Burns & Knox, 2011; Feryok & Oranje, 2015; Finch, 2010; Kiss, 2012; Svalberg, 2015). However, this relatively small body of work emerging from an ontological research perspective is expanding (Hiver, 2015).

2.4.4.1 Exemplar Study from SLTC Literature: Kiss (2012)

Kiss’ (2012) investigation of how teacher learning took place in an intensive post-graduate L2 teacher education course adopted complexity science as its research framework. He provided empirical evidence that learning in a teacher education course is a complex, non-linear experience. In his paper, Kiss (2012) used the six features of complexity science to explain the results of the student teacher journal reflections he analyzed. As a researcher working within this fourth ontology, Kiss mapped out the participating teachers’ cognitions through thematic analysis. The student teachers' cognitions spanned past, present, and future, and they also featured a variety of perspectives that included not only the learner but also the teacher and administrator. Kiss concluded that teacher learning can be viewed as dynamic, non-linear,

dependent on initial conditions (prior experiences), unpredictable, and chaotic, and that this perspective has important implications for teacher education programs, which (like many programs for students) are still generally structured around the idea of learning being a linear process.

2.5 Concluding Remarks on L2 Teacher Cognition Research

Research into L2 teacher cognition has changed over time, especially as related to ontological stances and methodological approaches. Four different research traditions have been identified: the individualist [1990 onwards] – a cognitivist ontological tradition grounded in teachers’ decisions and decision-making practices, thoughts, and beliefs, which are mainly analyzed through quantitative means. Then, the social era [1995 onwards] – a sociocognitive paradigm in which researchers analyze the wider surroundings and how the context shapes or informs thinking, analyzed mostly through qualitative means such as using diary studies. The third tradition is identified as sociohistorical [2000 onwards] – “thinking as a function of place and time, through interaction and negotiation with social and historical contexts” (Burns, Freeman, & Edwards, 2015, p. 589), which has been researched quite extensively through qualitative measures, including interviews and narrative inquiry. The fourth ontological tradition is the complex/chaotic systems [2010 onwards] – a dynamic and emergent system that involves the integration of multiple interconnected elements, and has been researched qualitatively through analysis of interactions. These four ontological research eras constitute an overall framework that helps to organize and define the various periods and methodological themes reflective of contemporary language teacher cognition literature. The four-component framework will be featured in the following study of L2 teachers’ cognitions of assessment. The research methods presented above – survey research (with a validation technique, factor analysis), FGIs, and semi-

structured interviews – will be used to collect data on L2 teachers’ cognitions of language assessment.

2.6 Topics and Contexts Studied

Language Teacher Cognition research is diverse in its subject matter and has been conducted throughout many different second language (L2) and foreign language (FL) education contexts. The most notable includes L2 teacher cognition about grammar teaching and grammatical terminology (Andrews, 1999; Borg, 1999). Work in L2 grammar teaching has developed our understandings of the way teachers teach grammar and of the thinking behind their practices (e.g., in explaining grammar and correcting learners’ grammatical errors.) Research focusing on teachers’ beliefs about formal grammar teaching instruction has taken place mostly in English as a Foreign Language (EFL) contexts. It has revealed that (1) teachers report that teaching grammar is important; (2) teachers’ views on grammar teaching were impacted by their previous language learning experiences; and, (3) teachers’ and students’ views of grammar teaching vary dramatically, which could affect the formal instruction teachers provide (Borg, 2003b). Additionally, these studies illuminate the complex nature of instructional decision-making processes in the formal instruction of grammar. Certain factors play a role in how language teachers approach the teaching of grammar – declarative knowledge, knowledge about the classroom environment, an understanding of their learners, and their own past histories learning/teaching grammar.

Some of the less researched content (subject matter) areas include pronunciation teaching (Baker, 2014; Burri, Chen, & Baker, 2017), reading instruction (Gilje, 2014), and writing (Cumming, 2003). Baker’s (2014) pronunciation study included an in-depth analysis of how teachers implement their beliefs “efficiently and successfully in the classroom to assist students

to achieve comprehensible pronunciation” (p. 138). The lack of empirical, classroom-based research on teacher cognitions and pronunciation teaching was undertaken with five experienced teachers in a North American IEP program by asking the following research questions: (1) What cognitions do experienced teachers have about techniques for teaching L2 pronunciation in their oral communication classes; and, (2) What do classroom observations and student questionnaires reveal about the teachers’ knowledge and practices concerning the techniques they use for teaching pronunciation? Baker’s findings revealed that teachers' knowledge of pronunciation techniques consisted mainly of controlled practice, and were limited to guided (semi-controlled) techniques (more extemporaneous practice opportunities were even less common). Additionally, the teachers held three basic beliefs about teaching pronunciation: (1) listening perception is essential for producing comprehensible speech, (2) kinesthetic/tactile practice is integral to phonological improvement, and (3) pronunciation instruction can sometimes be monotonous.

In addition to L2 teacher cognitions and the teaching of pronunciation, the subject matter of reading has also been examined. For instance, Gilje’s (2014) reading instruction study aimed to explore the knowledge, attitude, and beliefs which formed the basis of upper-primary EFL teachers’ reading-related materials and practices. Gilje asked the following questions: (1) What materials and practices do the 6th-grade EFL teachers employ in the teaching of reading; (2) What knowledge, attitudes, and beliefs form the basis of their choices; and, (3) What role does EFL teacher education play in this context? She conducted eight semi-structured interviews with 6th grade teachers and analyzed the data looking for trends and deviances across the interviews. Gilje found that the teachers based their teaching of reading on the textbook, and utilized various reading materials to varying extents. The teachers appeared to be heavily guided more by their textbooks than their intuition and routines (Gilje, 2014). The impact of formal teacher education

on reading instruction varied from teacher to teacher, and Gilje posited that pre-service and in-service teacher education programs could play important roles in helping future EFL teachers make and better understand their choices concerning reading materials and their instructional practices.

In addition to the body of research on L2 teacher cognitions and grammar, pronunciation, and reading instruction, there is work on L2 writing teacher cognition. Cumming (2003), Lipa and Harlin (1990), and Tsui (1996) represent three foundational studies which examined L2 teachers' beliefs and teaching practices. Cumming (2003) looked at L2 writing instructors' conceptualizations of the teaching of L2 writing; Lipa and Harlin (1990) reported on teachers' beliefs about process approaches to the teaching of writing; and Tsui (1996) examined L2 writing teachers' realizations regarding their integration of a process-oriented approach to their teaching of writing. Many of the studies that examine L2 teacher cognitions and L2 writing teachers teaching practices have found a gap between beliefs and instructional practices.

With the popularity of researching whether L2 teachers' beliefs and teaching practices match, Basturkmen (2012) decided to synthesize the published research that examined this phenomenon. Basturkmen's study adopted a meta-analysis approach from the quantitative/positivism paradigm, and used Interpretive Synthesis (Dixon-Woods et al., 2006). Interpretive Synthesis creates criteria for literature to be included, selects and finalizes the literature, and finally, tries to generate "concepts that have maximum explanatory value" (p. 284). Basturkmen discovered four main themes throughout the limited literature review (17 articles). First, there is a relationship between the context that the teacher works in and his/her stated beliefs. Second, the stated beliefs of more experienced teachers corresponded more highly with their classroom practices when compared to those who were less experienced. Third, there

is a correspondence between a teacher's stated beliefs and their planned aspects of teaching. These results are self-evident for Basturkmen, and therefore, she suggested that researchers who are interested in teacher cognition (particularly the relationship between beliefs and practices) should examine teachers' implicit beliefs along with what happens in the classroom incidentally. Such a focus, she believes, would broaden this research area. Finally, there is insufficient evidence to argue whether there is a correspondence between teachers' stated beliefs and the instructional methodologies they employ. In essence, "the review suggested that stated beliefs appear to be a more 'reliable guide to reality' (Pajares, 1992, p. 326) where experienced teachers (compared to new teachers) and planned aspects of teaching were involved" (p. 291). Basturkmen concludes with limitations of the study (the limited number of articles and the high number of Ph.D. dissertations) and implications for language teacher educators such as recording teachers and using unplanned actions as a basis for discussion and reflection.

In addition to the studies of particular language skills reviewed above, there are also teacher cognition studies that focus on language teacher identity (Kanno & Stuart, 2011) in the context of their beliefs and understandings of culture (Feryok & Oranje, 2015). Additionally, there are some studies that focus on the general process of language and language as pedagogical content (Breen et al., 2001); prior learning experiences (Farrell, 1999); language teaching practices (Basturkmen et al., 2004); and language learning processes (Macdonald et al., 2001).

As exemplified above, researchers who investigate L2 teacher cognition have examined many different content and subject matter areas, including grammar teaching and grammatical terminology, pronunciation, reading, and writing instruction; fewer studies have addressed L2 cognition and the teaching of vocabulary, speaking, and listening. Some studies that do not focus specifically on L2 teacher cognition and the teaching of language skills or curricula content,

identify general processes such as the knowledge growth and teacher identity (Burri, Chen, & Baker, 2017). Through these studies, language teacher cognition scholars have discovered that cognitions can be complex, ranging over many different subjects; they can be dynamic, changing over time and under different influences; and they can be systematic, forming unified and cohesive personal and practical theories (Feryok, 2010). Although not especially recent, Borg's (2009) discussion highlighted the findings of teacher cognition studies which have been found to be vital to our understanding of what we have learned about L2 teacher cognitions today, in that teachers' cognitions can be:

powerfully influenced by their own experiences as learners; these cognitions influence what and how teachers learn during teacher education; they act as a filter through which teachers interpret new information and experience; they may outweigh the effects of teacher education in influencing what teachers do in the classroom; they can be deep-rooted and resisted to change; they can exert a persistent long-term influence on teachers' instructional practices; they are, at the same time, not always reflected in what teachers do in the classroom; and they interact bi-directionally with experience. (p. 3)

Despite the aforementioned studies, one subject matter area that has been underexplored in the literature – but is a critical area of language teaching (and explained in the Introduction) – is the relationship between L2 teacher cognition and assessment, assessment practices, and language testing. Teachers' cognitions about assessment have been underexplored from a L2 teacher cognition perspective, but there is a fairly extensive literature on what is known as language assessment literacy (see below in Section 2.8), some of which deals with teacher cognitions. Based on what I believe to be a comprehensive review of 100 relevant teacher assessment literacy studies, there have been many empirical and conceptual studies from general education

that make connections between teachers' cognitions and assessment (e.g., Quilter & Gallini, 2000; Hill et al., 2010). However, from an L2 teacher cognition literature perspective, there have not been any empirical or conceptual studies conducted on connections between L2 teacher cognition, particularly EFL teachers, and language assessment (e.g., assessment practices).

2.7 Relationship Among Teaching, Learning, and Assessing

As discussed in the Introduction, language teachers have a dual role of teaching and assessing, which ultimately impacts student learning and motivation. Their assessment-related decisions serve to determine who will pass or fail a quiz, test, or course of study; whether the class is going well; and the degree of effectiveness of the teaching process (Harding & Kremmel, 2016).

Teachers will need to be able to design, carry out, and interpret a variety of different assessment approaches that have often been identified in the assessment literature as formative and summative assessments. *Formative assessments* are all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used to modify the teaching and learning activities in which they are engaged (Black & William, 1998). Additionally, in Black and Williams (1998) influential study *Inside the Black Box*, they explain assessment is formative when: (1) it is an integral part of the learning and teaching process; and (2) assessment evidence is actually used to modify teaching to meet the needs of pupils and improve learning. However, the integration of formative assessment as a teaching-learning-assessing model has been conceived differently with practicing teachers. Formative assessments are typically thought of by teachers – and different from scholars – as those along-the-way classroom tests and activities that teachers create to help them and their students understand how well students are learning what they are supposed to learn (Popham, 2011). Popham explains that “A number of commercial vendors describe their ‘interim tests,’ or their standardized tests

administered every few months, as incarnations of ‘formative assessment’”

(<http://www.edweek.org/ew/articles/2011/02/23/21popham.h30.html?t>, retrieved on August 10, 2018). Thus, there is an inaccurate conception of an instructional approach that could have innumerable benefits for teaching and learning (Birenbaum, 2014). For teachers to understand the instructional approach of formative assessment, Popham (2011) more accurately labels formative assessment as the ‘formative-assessment process’. His addition of the word, process, emphasizes that formative assessment is not a stand alone assessment, but is used in conjunction with the entire procedure of using assessments for the service of learning

In contrast to formative assessments, *summative assessments* are regarded by many educators as the test used to make evaluative judgments about a completed instructional sequence. Thus, summative assessments can range from large-scale standardized state examinations to classroom-based tests such as end-of-unit or chapter tests and final examinations. The purpose of summative assessments is to examine the extent to which knowledge (e.g., the material covered from the curriculum) has been acquired.

Formative and summative assessments are key components to the formative and summative assessment processes for the betterment of student learning and teachers’ pedagogy. The entire process involves decisions about when and what to assess and how to use the information elicited to make choices about adjustments in student’s learning or a teachers’ teaching. Popham (2011) explains,

When we employ phrases such as “a formative assessment” or “a summative assessment,” we are simply being sloppy with our language. Unfortunately, many educators truly believe formative assessment refers to particular kinds of tests that will—based on ample research evidence—improve kids’ learning. This simply is not so.

(<http://www.edweek.org/ew/articles/2011/02/23/21popham.h30.html?t>, retrieved on August 10, 2018).

To more fully illuminate how formative and summative assessments could be used as a *process* for learning, I will incorporate the recently used terms in the assessment literature: *Assessment-for-Learning* and *Assessment-of-Learning*. (I include the hyphenation and italics for ease of reading.) The definition of *Assessment-for-Learning* emerged from the first International Symposium on Assessment (2001) in which the term was defined as the following:

Assessment for Learning is formative assessment plus the deep involvement of learners in the assessment process. It is a process of both learners and teacher being engaged in seeing and interpreting evidence to figure out where learners are in their learning in relation to what has been taught, where they need to go next in their learning and how best to get there. The processes that support this work include having clear learning goals, co-constructing criteria around quality and success, engaging in all forms of feedback for learning (self-assessment, peer assessment, feedback from others), collecting evidence of learning and using information to guide the next learning steps. (as cited in, Davies, Busick, Herbst, & Sherman, 2014, p. 568)

Assessment-of-Learning and *Assessment-for-Learning* “are aligned with different outcome responses (p. 857)” coined by Bourke and Mentis (2013) as *feedback* and *feedforward*. *Feedback* is given to students after summative assessment, while *feedforward* is attributed to formative assessment. *Feedback*, explained by Bourke and Mentis (2013), is provided to students to show their achievements or failures on a task or knowledge of a particular area of content, and is seen as product-oriented. In contrast, *feedforward* is anticipatory in that it provides students with focused information that they can use in the next steps of the task, for future learning, or for

specific outcomes (Hattie & Timperley, 2017). Bourke and Mentis (2013) argue that *feedforward* is an approach that “all students benefit from” (p. 857). Thus, with *Assessment-of-Learning* and *Assessment-for-Learning*, language teachers need to be aware that the assessment practices that they choose will have an influence on students’ learning and motivation. Researchers have shown that when teachers use *Assessment-for-Learning*, students learn more and teaching becomes more effective (Andrade, 2013; McMillan, 2013). Although, as stated above, formative and summative assessments are used by the majority of scholars, Popham (2011), tallied the number of studies supporting *Assessment-for-Learning* at more than 4000 adding to the evidence that this is no longer a new way of thinking (Shepard, 2000).

Additionally, depending on factors such as the part of the world in which language teachers are operating (e.g., country), language being taught, program type, teacher background, and learner aspirations, language teachers need to understand that their instructional efforts take place within a broader societal context and curriculum. For example, the broader societal context and curriculum may influence or be influenced by high-stakes exams (e.g., district, state, or federal/national levels), or large-scale language standards (e.g., Common European Framework of Reference – CEFR).

External standards are an important part of a teacher’s professional life, and often play a role in the development of language curricula. One influential set of standards is the CEFR. The CEFR is meant to overcome communication issues among professionals working in the field of modern languages arising from the different educational systems in Europe (CEFR, 2018). The creators of the CEFR identified the communicative objectives and co-constructed different levels ranging from Level A1 (identified as “breakthrough”) to Level C2 (identified as “mastery”). From the CEFR, curricula and syllabi have been created and then distributed to teachers. The

CEFR is intended to serve as a set of external standards to help empower teachers (and, to a certain extent, learners) to set individualized goals, monitor and assess performance on learning tasks, and make decisions about the steps learners need to make for progress.

With the many different decisions language teachers must face in their day-to-day lives, there has been a growing concern among those involved in the outcomes of assessment decisions to assist in the development of teachers' language assessment literacy.

2.8 Assessment Literacy

Stiggins (1991) coined the term *assessment literacy* to refer to the knowledge, skills, and principles of sound assessment (Stiggins, 1991; 2002). From the academic field of language testing, Taylor (2009) took up Stiggins' (1991) definition and extended it. She conceptualized assessment literacy as a toolbox of competencies – the *know-how* – some practical and some theoretical, on why, when, and how to go about constructing, using, and interpreting a variety of assessment procedures. Although Taylor (2009) recognizes there are a wide variety of stakeholders of assessment (e.g., language teachers, administrators, policy makers, etc.), she lists the cognitive processes that underpin gaining assessment literacy (see Table 1) without clarifying differences among stakeholders.

Table 1 Assessment Literacy Involves..., adapted from Taylor (2009)

-
- | | |
|---|---|
| 1 | An understanding of the principles of sound assessment |
| 2 | The know-how required to assess learners effectively and maximize learning |
| 3 | The ability to identify and evaluate appropriate assessments for specific purposes |
| 4 | The ability to analyze empirical data to improve one's own instructional and assessment practices |
| 5 | The knowledge and understanding to interpret and apply assessment results in appropriate ways |
| 6 | The wisdom to be able to integrate assessment and its outcomes into the overall pedagogic/decision-making process |
-

These six factors (Taylor, 2009) all begin with a person's faculties, such as understanding, ability, knowledge, and wisdom. There is thus an underlying assumption that if one learns the principles, practices, and skills of assessment, then they may be characterized as being assessment literate.

2.8.1 Language Assessment Literacy

The term *language assessment literacy* is derived from the generic assessment literacy construct operationalized above. Language assessment literacy is another layer concerned with all aspects of assessment literacy plus issues that surround language, such as what the nature of language is, what the structural and functional conceptualizations of language are, and how languages (and, in particular, second languages) are learned or acquired. Inbar-Lourie (2008) explains that “formed on the assessment literacy knowledge base LAL can be said to constitute a unique complex entity” (p. 393). When language testing specialists refer to language assessment literacy for language teachers, they often refer to the level of teachers' engagement with testing and assessment issues to evaluate, select, and use appropriate instruments and techniques for specific purposes and contexts (Inbar-Lourie, 2008; Malone, 2008; Stiggins, 2002).

Language assessment literacy for language teachers is fundamental (Popham, 2009) and should form an integral part of teachers' professional development (see Section 2.7 for relationships among teaching, learning, and assessing). To identify the assessment training needs of language teachers, Fulcher's (2012) seminal article attempted to empirically uncover what language teachers value about assessment/testing. After conducting a survey study with 278 language teachers, he was able to expand the language assessment literacy construct for language teachers to include:

The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-

scale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals. (Fulcher, 2012, p. 125)

Fulcher (2012) provides a visual synopsis of an updated definition as follows:

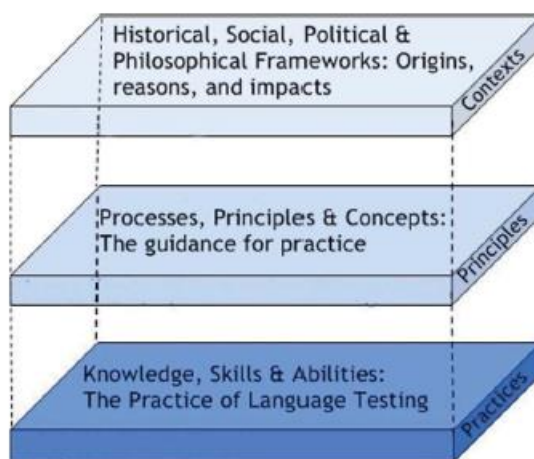


Figure 6 Expanded definition of LAL (Fulcher, 2012, p. 126)

Here, in Figure 6, Fulcher (2012) conceptualizes the properties and processes of assessment on a three-tiered structure with each of the three tiers being of equal size and shape. Although he does not explicitly feature the concept of culture in his definition or framework, I conceptualize it as a part of the Contexts level, which is located at the top tier of Figure 6. Fulcher's (2012) tripartite model has content that is similar to Taylor's (2009) view, outlined above in Table 1. However, his expansion is different from Taylor's (2009) in two key ways. First, he acknowledges that not all components of Taylor's model will be required for each stakeholder of language assessment.

Second, he shows that language teachers' prior and current experience in language teaching influences how they understand components of language assessment literacy (i.e., tier 2 – 'Process, Principles, & Concepts: The guidance for practice'). Fulcher's (2012) working definition of language assessment literacy above, with his tripartite model, has propelled further research into discovering the components of language assessment literacy that might be required for different stakeholder groups (Taylor, 2013; Harding & Kremmel, 2016).

2.8.2 Developing Language Assessment Literacy

Other research in language assessment literacy has attempted to describe how assessment literacy might develop. An approach adopted in the fields of scientific and mathematical literacy education is the rejection of a dichotomy of "literacy" or "illiteracy" in preference for viewing degrees of literacy falling along a continuum. To provide a conceptual understanding of how the assessment literacy construct can grow over time, Pill and Harding (2013) took up the different levels of achieving scientific literacy (Bybee, 1997; Kaiser & Willander, 2005). There are five stages of literacy identified and described below: illiteracy, nominal literacy, functional literacy, procedural and conceptual literacy, and multidimensional literacy (see Table 2, based on Bybee, 1997; Kaiser & Willander, 2005; as cited in Pill & Harding, 2013, p. 383). Pill and Harding (2013) adapted these descriptors for language assessment literacy content.

Table 2 Levels of LAL, adapted from Bybee (1997), cited from Pill & Harding (2013), p. 383.

Illiteracy	Ignorance of language assessment concepts and methods
Nominal literacy	Understanding that a specific term relates to assessment, but may indicate a misconception
Functional literacy	Sound understanding of basic terms and concepts
Procedural and conceptual literacy	Understanding central concepts of the field, and using knowledge in practice
Multidimensional literacy	Knowledge extending beyond ordinary concepts including

Pill and Harding (2013) focused on the cognitive dimensions at each stage of gaining assessment literacy and placed different aspects of assessment literacy on a continuum or a progressive scale. There are two limitations to the assessment literacy continuum. First, the descriptions at each level are not consistent with each other, in which some levels overemphasize procedural / theoretical knowledge (i.e., illiteracy, nominal literacy, functional, and multidimensional) while one level refers sporadically to practice (i.e., procedural and conceptual literacy). Secondly, the continuum is not clear regarding what level a classroom teacher, government official, administrator, or any other stakeholder of language assessment needs to achieve to become assessment literate.

Taylor (2009) and Fulcher (2012), in contrast to Pill and Harding (2013), do not necessarily conceptualize achieving literacy as a step-by-step process (e.g., moving from being illiterate to having nominal literacy to having functional literacy to procedural and conceptual literacy, etc.). Instead, Taylor (2009) and Fulcher (2012) frame the uncovering of the ‘what’ of assessment to be assessment literate, while Pill and Harding identify ‘how’ assessment literacy develops for all stakeholders.

2.8.3 Identifying the Needs of Stakeholders in Language Assessment

A stakeholder in language assessment has a vested interest in the results of language testing or assessment and often makes decisions based on scores for various purposes. As stated above in 2.6, Taylor (2009) identified various groups of people including language teachers, administrators, policy makers, and the general public as stakeholders. Fulcher (2012) began to empirically discern the needs of the stakeholder group of language teachers. To hypothesize and

call for researchers to identify different needs for different stakeholder groups, Taylor (2013) created diagrams with eight axes representing dimensions of assessment literacy (see Figure 7).

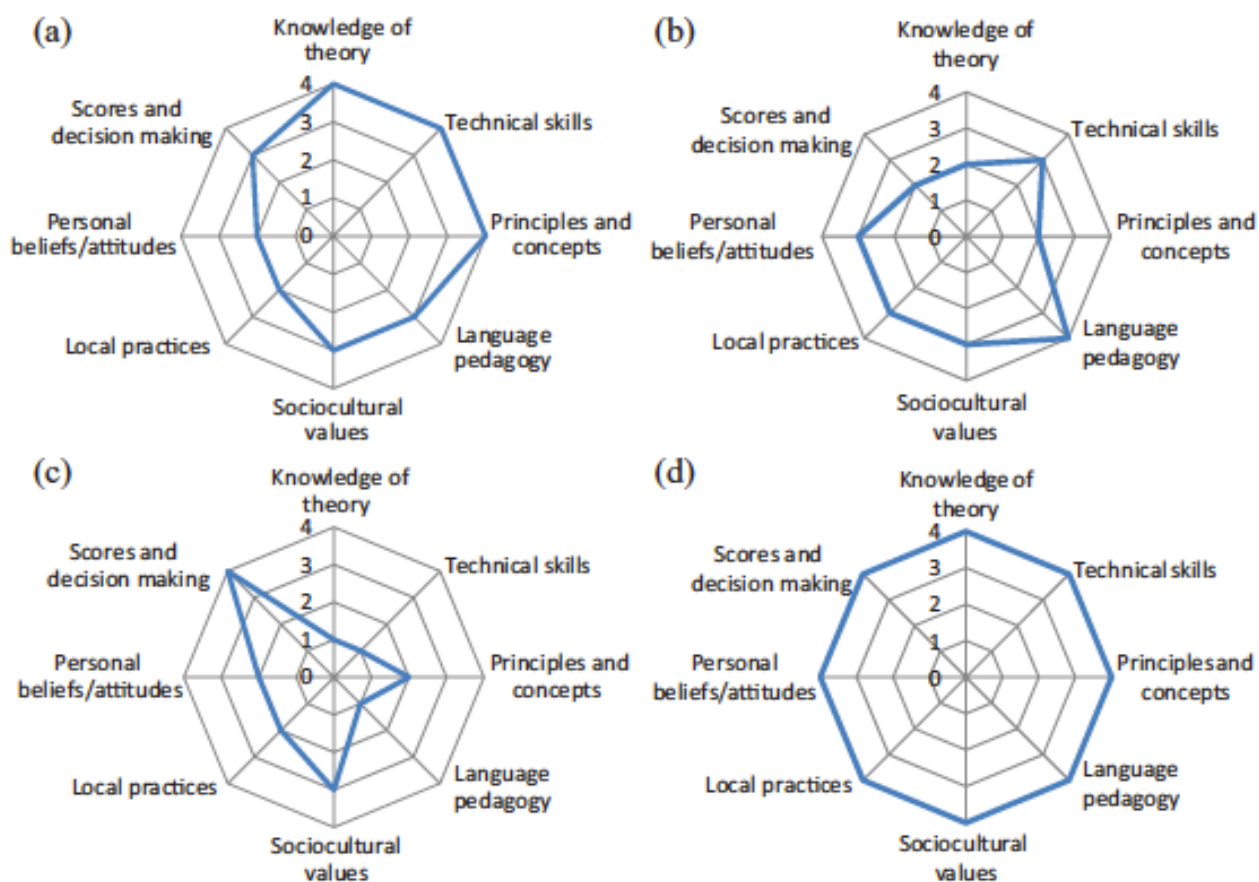


Figure 7 Levels of AL/LAL differentiated according to stakeholder constituency

The labeled dimensions on the eight axes are: knowledge of theory, technical skills, principles and concepts, language pedagogy, sociocultural values, local practices, personal beliefs/attitudes and scores and decision making. These categories are gleaned from the discussion of possible assessment literacy/language assessment literacy components across various papers in *Language Testing* [2013], while the values (i.e., 0-4) are hypothesized according to the different stages of literacy development suggested by Pill and Harding (2013). The dimensions of assessment literacy are similar in content with the conceptualizations of assessment literacy of Taylor (2009)

and Fulcher (2012). Figure 8 presents the levels of assessment literacy / language assessment literacy for classroom teachers, specifically.

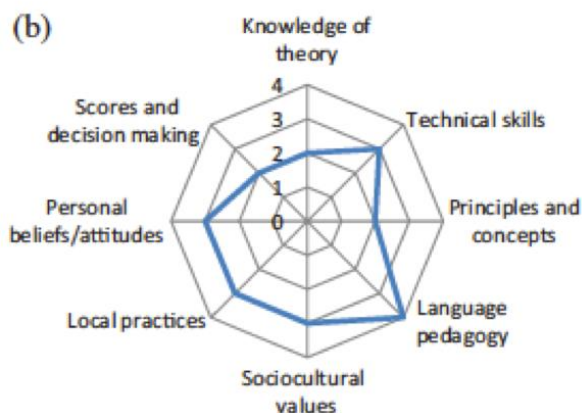


Figure 8 Levels of AL/LAL for classroom teachers

I will briefly explain the way to “read” the spider web images depicting the needs of classroom teachers, because teachers are the focal participants for the dissertation. [Each level’s definitions reported by Pill and Harding (2013) are displayed in Table 2 above.] A language teacher needs to have *functional literacy* (level 2) – a sound understanding – of the terms and concepts about knowledge of assessment theory, principles and concepts of assessment, and scores and decision-making. Additionally, language teachers need to have *procedural or conceptual literacy* (level 3) – understanding central concepts of the field, and using knowledge in practice – regarding technical skills of language assessment, local practices, and personal beliefs/attitudes. Finally, language teachers should have *multidimensional literacy* (level 4) – knowledge extending beyond ordinary concepts including philosophical, historical and social dimensions of assessment – about sociocultural values and language pedagogy.

Kremmel and Harding (*forthcoming*) took up Taylor’s (2013) call for researchers to identify different needs for different stakeholder groups. They proposed the following four research questions: (1) What levels of LAL do different stakeholder groups think they need; (2)

What levels of LAL do different stakeholder groups think other stakeholder groups need; (3) What levels of LAL do different stakeholder groups currently believe they have; and (4) What do different stakeholder groups believe they need in order to develop LAL to the target they have set? To address these questions, Kremmel and Harding created a Language Assessment Literacy survey and distributed it online to over 1000 participants. The survey was divided into two sections. The first section asked the following question: How knowledgeable do people in your chosen group/profession need to be about each aspect of language assessment below? The second section asked the following question: How skilled do people in your chosen group/profession need to be about each aspect of language assessment below? The survey was composed of 72 closed-item responses, for which a 0 (not knowledgeable at all, not skilled at all) to 4 (extremely knowledgeable, extremely skilled) Likert scale was used. The Language Assessment Literacy survey is one of many questionnaires that seeks to identify the assessment needs of teachers, particularly language teachers.

2.9 Methodology in LAL Research – Questionnaires and Validation

Questionnaire research has been the main method assessment literacy scholars have used to measure teachers' cognitions on and/or mastery of assessment concepts. Questionnaire research began in the context of teachers' language assessment literacy after the creation of the Standards for Teacher Competences in 1990 (henceforth, *Standards*). The 1990 assessment *Standards* from educational assessment (AFT, NCME, & NE, 1990) were a joint effort between the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. (See Table 3 for *Standards*).

Table 3 Standards for Teacher Competence (AFT, NCME, & NE, 1990)

Standard	Description
1	Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2	Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3	The teacher should be skilled in administering, scoring and interpreting the results of both externally produced and teacher-produced assessment methods.
4	Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
5	Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments.
6	Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
7	Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

All seven *Standards* acknowledge and address classroom-based assessments with instructional goals and objectives. Each *Standard* can apply to teachers' development of assessment practices, and were seen to be helpful for teacher educators to support teachers' development of assessment literacy (Mertler, 2003).

Research instruments were then developed from the *Standards* to investigate teachers' knowledge base needed to achieve assessment literacy. The two most-used questionnaires were the Teacher Assessment Literacy Questionnaire (TALQ; Plake, Impara, & Fager, 1993) and the Classroom Assessment Literacy Inventory (CALI; Mertler & Campbell, 2005). These objective questionnaires were tests that intended to measure what teachers knew about the prescribed *Standard* competencies and to identify strengths and weaknesses in assessment literacy. CALI consisted of five real-life scenarios with seven questions pertaining to each standard. These questions were written to be meaningful to teachers' actual practices. However, Fulcher (2012)

points out that Plake and Impara's (1996) study, which used CALI, showed that "there was a correlation between [teachers'] experience and score, but it was not possible to detect the assessment literacy needs of teachers" (p. 117). Thus, over the next generation, scholars in educational assessment created their own questionnaires (e.g., McMillan & Schumacher, 2001; Hasselgreen, Carlsen, & Helness, 2004; Huhta, Hirvalä, & Banerjee, 2005) to try to tap into the assessment needs of language teachers so that purposeful teacher development programs can build language teachers' assessment literacy.

In doing so, questionnaires were adapted and/or created to gauge the assessment needs of teachers, and the factor analysis technique became more common throughout articles to demonstrate a questionnaire's validity. One example of a study in assessment literacy that utilized a questionnaire and then used a factor analysis to contribute claims to its validity is DeLuca and Klinger (2010). They utilized a questionnaire and administered it to 288 pre-service teacher candidates in a teacher preparation program in Ontario, Canada. The questionnaire was adapted from McMillan and Schumacher (2001) and addressed pre-service teachers' confidence levels in educational assessment knowledge. The items addressed the three main categories of practice, theory, and philosophy within two major paradigms – *Assessment-of-Learning* and *Assessment-for-learning*. In total, the 57-item questionnaire consisted of three sections: Section A contained demographic information, section B included items that ranked confidence levels regarding specific assessment constructs within *Assessment-of-Learning* and *Assessment-for-Learning*, and section C was a checklist in which DeLuca and Klinger asked the respondents where they had learned their assessment knowledge, based on the items in B about which they felt they were 'very confident.' After the questionnaires were collected and values were placed in SPSS, DeLuca and Klinger calculated descriptive statistics and conducted a factor analysis and

an analysis of variance (ANOVA) to analyze the responses. They remarked, “In general findings support the need for direct instruction in assessment with specific topics identified (e.g., reporting achievement, modifying assessments, developing constructed-response items, item reliability, validity, articulating a philosophy of assessment, etc.) as important to developing teacher assessment literacy” (p. 419). This article highlights the importance of identifying the assessment needs of teacher candidates with the research methods of questionnaire design and validation, so as to better provide a foundation for understanding initial assessment literacy development.

Fulcher’s (2012) questionnaire study attempted to identify the language assessment training needs of language teachers empirically, gathering online survey responses of 278 teachers involved in language teaching in a variety of contexts. Fulcher (2012) wanted to find out the assessment needs of language teachers so that he could inform the development and writing of a new language testing manual. Through development, piloting, and analysis (with a factor analysis technique), Fulcher (2012) found that the theoretical concepts of language assessment and testing should be presented to language teachers within the context of practical test construction, “using the test development cycle as the scaffold and introducing principles and core terminological knowledge along the way rather than merely introducing them a decontextualized component of LAL” (Harding & Kremmel, 2016, p. 419). An implementation of the approach can be seen in Fulcher’s (2010) textbook *Practical Language Testing*.

As stated previously, the Language Assessment Literacy survey (Kremmel & Harding, *forthcoming*) is one of the more recent questionnaires that tries to uncover the training needs of language teachers. Specifically, Kremmel and Harding wanted to investigate different stakeholder (i.e., language teachers, administrators, testing professionals, policy makers)

perceptions of needs and deficiencies of language assessment/testing. As surveys and questionnaires are the driving research instruments to understand teachers' cognitions on language assessment literacy, I take up Kremmel and Harding's survey for the dissertation for three reasons. First, the content matter of the survey is written for stakeholders in *language assessment*, as opposed to many of the other questionnaires on assessment literacy [e.g., Classroom Assessment Literacy Inventory (Mertler & Campbell, 2005)]. Second, the Language Assessment Literacy survey contains items that address specific skills and knowledge about assessment/testing practices that professionals and academics from the language testing community find to be important. This reason is particularly relevant because Kremmel and Harding's survey could be a valuable tool that could be used in a variety of specific contexts to obtain useful information for teacher educators and assessment scholars to inform both theory and practice. Third, the Language Assessment Literacy survey accounts for the depth of knowledge of language assessment content matter gleaned from the research literature on language assessment literacy (i.e., Pill & Harding, 2013). Because Kremmel and Harding's survey purports to be based on a theoretical framework (i.e., Taylor 2013), I incorporate the factor analysis method common in studies on language assessment literacy to provide evidence for the construct validity of the questionnaire. Now, I examine how the fields of educational assessment, teacher education, and language assessment literacy have expanded the scope of research of assessment literacy beyond the knowledge and skill base of assessment.

2.9.1 Beyond Assessment Needs for Teachers

Xu and Brown (2016) published an article titled, "Teacher Assessment Literacy in Practice: A reconceptualization" in the journal *Teacher and Teacher Education*. This article aimed to reconceptualize teacher assessment literacy by connecting the two fields of research: educational

assessment and teacher education. They have conceptualized assessment literacy for classroom language teachers as a construct that is more than its knowledge base; they argue that, even though the knowledge base is a necessary condition for literacy (Fulcher, 2012), *teachers' assessment literacy* has a different nature/quality, because the principles of assessment are decontextualized in classrooms and “are not ready-made solutions to problems that arise within complex and diverse classroom assessment scenarios” (Xu & Brown, 2016, p. 156). Insights from assessment literacy as a social practice situated in multiple discourses were an impetus for Xu and Brown (2016) to reconceptualize assessment literacy as something constantly in flux with the micro (e.g., schools) and the macro (e.g., sociocultural) institutional contexts. Teacher cognition researchers eventually settled upon the term *pedagogical content knowledge* (i.e., subject-specific instructional techniques) to refer to such practices, and Xu and Brown (2016) have incorporated this cognition-related terminology in their guiding conceptual framework below (see Figure 9) in the bottom row labeled *The Knowledge Base*. Instead of understanding assessment literacy as a literacy or literacies, Xu and Brown (2016) coined the term TALiP (Teacher Assessment Literacy in Practice) and created a conceptual framework (see Figure 9), which contains seven main areas that interact within and through each other: the knowledge base, teachers' interpretive and guiding framework, teacher conceptions of assessment, macro socio-cultural and micro-institutional contexts, teacher assessment literacy in practice (the core of the construct), teacher learning, and assessor identity (re)construction.

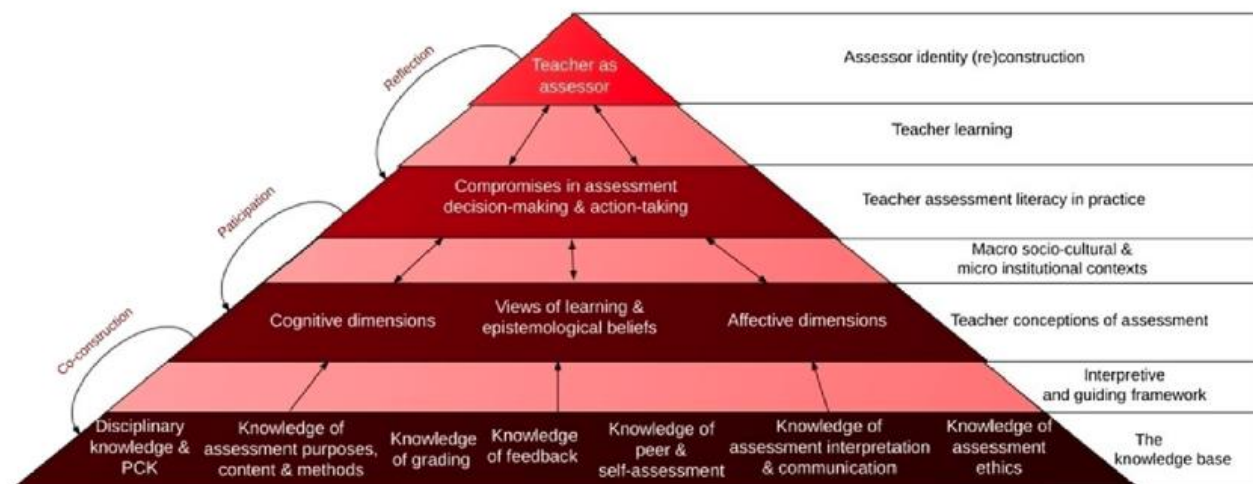


Figure 9 A conceptual framework of teacher assessment literacy in practice

Interestingly, the way to read this conceptual framework, whether top-down or bottom-up, depends on the positionality of the reader in relation to the academic concepts of assessment or teacher education. If a reader examining this framework is from an assessment-focused perspective (bottom up) or a teacher educator perspective (top down), they could read it and understand it differently. From a teacher educator's perspective, a teacher has multiple, fluid, and dynamic identities which interact within and through different contexts at all moments in time (Beauchamp & Thomas, 2009). A teacher educator could use Xu and Brown's (2016) diagram to understand how teachers' assessment identity develops within the micro (e.g., school) and macro (e.g., sociocultural) institutional contexts.

Xu and Brown (2016) come from the assessment paradigm and bring their knowledge of assessment literacy research to the teacher educator community. Their framework, as well as Fulcher's (2012) described above position the subject matter or *Knowledge Base* of assessment as the pyramid's base. The knowledge base is meant to support the teachers' identity as an assessor, which may be located at the top of the pyramid. Xu and Brown (2016) explain that,

At the bottom of this pyramid is the knowledge base, which is the basis of all other components, since clarity of the knowledge that both pre- and in-service teachers need for effective assessment practice is essential (Maclellan, 2004). Without the knowledge base, there would be no standards or criteria by which the appropriateness of assessment practice could be evaluated, potentially causing failed outcomes for teachers and students (Fulcher, 2012). (p. 155)

Perhaps, it is the labels on the chart / full diagram that also that might need to be reconceptualized more for teacher educators' perspective. The term *pedagogical content knowledge about assessment* (PCKAA) would be more consistent with acronyms used in the teacher education literature. PCKAA is having access to procedural knowledge, knowing how to apply knowledge about assessment (the base of Xu and Brown's [2016] chart) in relevant ways as part of language instruction – both inside and outside of the classroom context. PCKAA places the teacher as an agent of an action, using the knowledge of assessment.

The TALiP framework is cyclical in nature and has multidimensional flows, which shows that if one aspect of the model changes, through tensions, so does the rest of the model. These tensions have been discussed above as the challenges classroom teachers may have to face within the process of becoming assessment literate. Xu and Brown's (2016) conceptualization of assessment literacy emphasize the affective-volition challenges that teacher cognitions researchers have discussed as important factors for teachers' development. "Both teachers' reflective practice and participation in community-based assessment activities will generate affordances for teachers to (re)construct their identities as assessors" (Xu & Brown, 2016, p. 159).

2.10 Methodology in LAL Research – Surveys/Questionnaires + Interviews

Insights from assessment literacy (e.g., Xu & Brown, 2016) as a social practice situated in multiple discourses has provided the language assessment literacy research community more opportunity to use different methods to investigate teachers' language assessment literacy. Most studies use questionnaires plus focus group interviews (FGIs) and/or semi-structured interviews. For example, Lam's (2015) study incorporated a questionnaire, focus group, and semi-structured interviews. Lam's (2015) study was framed by the following inquiry, "What is the overall language assessment training landscape in Hong Kong and how does it influence the development of pre-service teachers' [language assessment literacy] LAL in a context of assessment reforms?" (p. 174). To gauge an overall understanding of the assessment landscape, Lam utilized surveys and examined teacher-training programs in five teacher education institutes. Then, Lam asked, "What are instructors' and pre-service teachers' perceptions of how course-based language assessment training promotes LAL?" (p. 176). The methods used included focus group interviews, individual/semi-structured interviews, and document analysis. Nine instructors and 40 pre-service teachers studying in the final year of a teacher education program took part in the study. "Findings indicate that language assessment training in Hong Kong remains inadequate and selected language assessment courses are still unable to bridge the theory-practice gap within the assessment reform context" (p. 169). One of Lam's most fruitful findings lies in the fact that the study "enriches scholars' understanding of the relationship between language assessment training and development of LAL within a context of assessment reforms" (p. 190). Lam's (2015) study contributed to the field by triangulating with a variety of research methodologies.

2.11 (Language) Assessment Literacy Conclusion

Assessment is a process of gathering information about a person's or a group of people's abilities, making inferences about their abilities, and then using these inferences to make decisions. This process can be accomplished through a variety of methods and perspectives, including *Assessment-of-learning* and *Assessment-for-learning*. The core processes of assessment do not change – inferences are still being made for decision-making purposes. For language assessment specifically, administrators, students, and the general public expect classroom teachers to be well versed in current theories and research findings regarding various facets of language knowledge and use. This aspect of a teacher's professional life has language testing/assessment specialists calling for a growing interest to raise the assessment literacy of classroom teachers.

In this chapter I presented various conceptualizations of language assessment literacy and research methods from educational assessment and language testing that investigate the assessment literacy construct. Taylor (2009; 2013), Fulcher (2012), Pill and Harding (2013) are some of the most influential in the language testing literature. Both Taylor's (2009) and Fulcher's (2012) conceptualizations present the 'what' of assessment that is necessary to be assessment literate; Pill and Harding (2013) discuss 'how' assessment literacy evolves. Taylor's (2013) dimensions address the question of whether a classroom language teacher has similar or different needs from other stakeholder groups in assessment. Her conceptualization has been taken up by Kremmel and Harding in their Language Assessment Literacy survey that I use for the dissertation. To extend beyond the knowledge base of language assessment with the research method of questionnaires, I looked to (and took up) Xu and Brown's (2016) TALiP conceptualization, which shows there is a relationship among teachers' cognitions surrounding,

feelings about, and activities of assessing, uncovered by the use of focus-group and semi-structured interviews. Thus, advancement of assessment literacy for teachers deals with not only the knowledge base of assessment practices and the sociocultural, historical, and cultural situated practices of assessment, but also the cognitive and emotional dimensions of teachers' inner lives when they are confronted with assessment decisions. In essence, I will identify the assessment needs of language teachers, and examine the cognitive-affect- activity dimensions of language teachers' lives as exemplified in Xu and Brown (2016). The specific methodological approach I take to research language teachers' cognitions about language assessment literacy comes from three of the ontological research generations in L2 teacher cognition research: individual, social, sociohistorical (Burns, Freeman, & Edwards, 2015).

2.12 Expanding the Research Domains of LTC and LAL

To concisely organize methodologies used in studies on language teacher assessment literacy, I have mapped them onto the ontological research traditions identified by Burns et al. (2015). In doing so, I identified the conceptual unit of study (e.g., decisions, thoughts beliefs about assessment) that each article in the language teacher assessment literacy literature aimed to analyze. Then I identified the prevailing research methodology used in each study. Thereafter, I informally compared what and how research was conducted in the assessment literacy literature to the teacher cognition research paradigms in order to identify the direction I will take to expand the language teacher cognition research agenda with assessment as its subject matter.¹ Of the twelve empirical published research studies since 1995 that focused on language teachers'

¹ For future purposes, I will calculate inter-rater reliability to show that the categorization is indeed accurate.

language assessment literacy, the majority fall within the individualist ontological paradigm ($N=8$), while five studies fall into the social ontological tradition; no studies are exemplified in the sociohistorical or chaos/complex generations of research (see Table 4).

Table 4 Research on Language Teachers' Language Assessment Literacy

Ontological Generation	Study
Individualist	DeLuca & Klinger (2010); Falvey & Cheng (1995); Fulcher (2012); Hakim (2015); Jin (2010); Volante & Fazio (2007); Xu & Liu (2009); Zolfaghari & Ashraf (2015)
Social	Hill & McNamara (2011); Koh (2011); Lam (2015); O'Loughlin (2006); Scarino (2013)
Sociohistorical	None
Chaos/Complex	None

The sections of the dissertation to follow will continue the traditions of the individualist and social ontological traditions, but will also expand the analysis of the relationship between L2 teacher cognitions and language assessment literacy to include the sociohistorical research tradition. More specifically, I will continue the tradition of the individual research paradigm by asking the following questions:

- 1) To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy?
 - A. What assessment skills and knowledge do Uzbekistan EFL teachers believe teachers need to possess?
 - B. Is the theoretical basis for the survey (Taylor's 2013 framework of LAL) faithfully implemented in the survey?
 - C. What is the factor structure present in the Uzbekistan EFL teachers' responses?

The first overarching research question and subquestions will provide a glimpse into the language assessment literacy of Uzbekistan EFL teachers. At the same time, the overarching

question will look at the extent to which the Language Assessment Literacy survey provides valid and actionable information about teachers' LAL in the Uzbekistan EFL context. Also, I will continue the tradition of the social research paradigm by asking the following questions:

- 2) How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students?
 - A. What do Uzbekistan EFL university teachers report they do to assess their students (i.e., assessment tasks and scoring procedures)?
 - B. What are their cognitions surrounding those assessment practices?
 - C. How do their cognitions about assessment shape how they assign scores to their students' work?

The broader research question and subquestions will examine the conceptual changes in thinking from a sociocognitive perspective by showing how the micro-institutional and macro-sociocultural contexts, both internal to the L2 teacher and external to the social setting, shapes or informs thinking about assessment (Burns, Freeman, & Edwards, 2015, p. 591). I will expand the L2 teacher cognition research agenda on assessment literacy to include the sociohistorical research generation, and will ask the following questions:

- 3) What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students?
 - A. What are the macro-environmental factors Uzbekistan EFL university teachers report?
 - B. What are their cognitions surrounding these factors?
 - C. How do the reported factors and cognitions shape Uzbekistan EFL university

teachers' assessment practices?

These questions will apply a more eclectic view of the language teacher mind, which reflects Borg's (2006) view of L2 teacher cognition as an integration of sources of knowledge that includes schooling, professional coursework, contextual factors, and classroom practices (see Figure 4). I will not only examine the participating L2 teachers' beliefs about assessment but will also analyze how micro-institutional and macro-sociocultural aspects of their lives as language teachers (including past, present, and future aspects) may be impacting their assessment practices. Aspects of the chaos/complex ontological research generation will contribute to the analysis, but not be at the forefront of the study.

2.13 Conclusion

In this chapter, I discussed how research into L2 teacher cognition has changed over time, especially as related to ontological stances and methodological approaches. Four different research eras have been identified: the individualist [1990 onwards] – a cognitivist ontological tradition grounded in teachers' decisions and decision-making practices, thoughts, and beliefs, which are mainly analyzed through quantitative means. Then, the social era [1995 onwards] – a sociocognitive paradigm in which researchers analyze the wider surroundings and how the context shapes or informs thinking, analyzed mostly through qualitative means such as using diary studies. The third era is identified as sociohistorical [2000 onwards] – “thinking as a function of place and time, through interaction and negotiation with social and historical contexts” (Burns, Freeman, & Edwards, 2015, p. 589), which has been researched quite extensively through qualitative measures, including interviews and narrative inquiry. The fourth ontological era is the complex/chaotic systems [2010 onwards] – a dynamic and emergent system that involves the integration of multiple interconnected elements, and has been

researched qualitatively through analysis of interactions. These four ontological research traditions present an overall framework that discerns the various periods and methodological themes throughout the language teacher cognition literature, and will be integrated into the study on L2 teachers' cognitions of assessment.

In this chapter I presented various conceptualizations of language assessment literacy [i.e., Taylor (2009; 2013); Fulcher (2012); Pill & Harding (2013)]. The relationship between the 'what' of assessment literacy and 'how' assessment literacy develops had been conceptualized in Taylor's (2013) dimensions. To empirically validate Taylor's (2013) dimensions, Kremmel and Harding developed and used the Language Assessment Literacy survey. Their survey will be integrated into the study on L2 teachers' cognitions on assessment to provide valid and actionable information about teachers' language assessment literacy in the Uzbekistan EFL context. I also presented Xu and Brown's (2016) TALiP conceptualization, which shows there is a relationship among teachers' cognitions surrounding, feelings about, and activities of assessing. For the study of L2 teacher cognitions on assessment, I will identify the assessment needs of language teachers, and examine the cognitive-affect- activity dimensions of language teachers' lives as exemplified in Xu and Brown (2016). Finally, I discussed the research methods (i.e., surveys, questionnaires, interviews) used in the academic fields of educational assessment and language testing to investigate the assessment literacy construct. Thus, advancement of assessment literacy for teachers deals with not only the knowledge base of assessment practices and the sociocultural, historical, and cultural situated practices of assessment, but also the cognitive and emotional dimensions of teachers' inner lives when they are confronted with assessment decisions.

3 SOCIOCULTURAL CONSIDERATIONS – STUDY BACKGROUND

In this chapter, I will discuss the sociocultural background to the setting of the study. First, I will explain why I chose Uzbekistan. Then, I will show how Uzbekistan's history, culture, and politics have influenced the development of English language teaching throughout the country. Moreover, I will discuss Uzbekistan's relationship with the United States, which has been a key factor to the development and implementation of the study. Finally, I will explain how the government of Uzbekistan has had an engaged role in the recruitment of participants.

3.1 Selecting Uzbekistan

Researchers who wish to investigate the language assessment literacy needs of teachers expand their research beyond the knowledge base of language assessment/testing to encompass the context, at both the institutional and larger macro sociocultural levels that the language teacher is a part of. A specific cultural context I have been interested in is Central Asia, a region of the world with a rich history and culture that has had a major impact on western countries in terms of religion, politics, and economics. Within Central Asia, I have been particularly interested in Uzbekistan because of its culture, political history, and language planning and policy, which impacted the people of Uzbekistan and the professional lives of language teachers. Additionally, I had an opportunity to conduct teacher development in Uzbekistan and wanted to understand what teachers there know and need to know about assessment. My personal interest and work opportunities in Uzbekistan helped me to select Uzbekistan as a site for research.

3.2 History and Culture of Uzbekistan

At the heart of Central Asia is the country of Uzbekistan. It is a country that is identified by Turks as the melting pot of Central Asia. Uzbeks often compare their identity to that of the United States, in which people from various cultures come together and co-exist. Islam Karimov,

who was the first President of Uzbekistan after the dissolution of the Soviet Union in 1991, explained that “if somebody wants to understand who the Uzbeks are, if somebody wants to comprehend all the power, might, justice and unlimited abilities of the Uzbek people, their contribution to the global development, their belief in future, he should recall the image of Amir Temur” (inscription on the wall at the Islam Karimov museum). Amir Temur (1370-1405), also known as Tamerlane in the west, was a military leader of great magnitude who has been equated to Alexander the Great. Amir Temur brought people together from all over Central Asia through his conquests, and he is a historical reference that the people of Uzbekistan admire. However, since Temur’s reign, Uzbekistan’s people have been a part of numerous battles, wars, and oppressions and this participation continued into the late 20th century with the Russian/Soviet occupation of Uzbekistan from 1865 to 1991.

The decades under Russian (1865-1922) and Soviet rule (1922-1991) were identified by a government minister in Uzbekistan as a period of “revolution, oppression, massive disruptions, and colonial rule” (U. Azizov, personal communication, July 15, 2017). In 1865, Russians took over Tashkent, the capital city of Uzbekistan, and made it the capital of a newly created province called Turkestan. The area of Turkestan covered much of Central and East Asia including modern-day regions of Afghanistan, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Xinjiang autonomous region of China. The Russian takeover of Central Asia not only influenced the political climate of the country but also molded the cultural milieu of Central Asia to introduce more Russian influence. For example, the Muslim faith waned as the new Communist rulers in 1918 closed mosques across the Turkestan province and persecuted Muslim clergy as part of a secularization campaign. Then, in the early 1930s, the Soviet leader Stalin purged “independent-minded Uzbek leaders” (retrieved from, <https://www.bbc.com/news/world-asia->

[16218972](#) on August 3rd, 2018) and replaced them in the government with Moscow loyalists.

This replacement had a major effect on the languages that were used in Uzbekistan's government and with the people in the cities and towns. The Russian language became the dominant means of communication with government and business. Furthermore, the Russian language, not the Uzbek language, began to be used as the means of instruction in higher education institutions across the country.

In 1989 Islam Karimov became leader of the Uzbek Communist Party after his predecessor failed to quell inter-ethnic disputes in the Fergana Valley. The colonial rule finally ceased in 1991 with the fall of the Soviet Union, and Karimov was voted in as the first President of Uzbekistan. Karimov (1999) emphasized that "Independence gave the Uzbek people a possibility to revalue its historical heritage, to revitalize the feeling of national self-respect, its cultural traditions, faith, language and spirituality. Independence became a new pulse in the development of the Uzbek national mentality, strengthening the feeling of patriotism and love to the motherland" (inscription on the wall at the Islam Karimov museum). The love of the motherland was strong in Uzbekistan. What was not anticipated by government officials was a revitalization of all aspects of Uzbek society, including its education system, which presented a language policy concern. The government of Uzbekistan had to choose whether they would want to revive the Uzbek language across schools and social systems and in day-to-day interactions, or to continue using Russian as the language of the government. The language of choice was an important decision to make for the Uzbek government because it privileged dominant groups and made a stance about how they want to view the world and be seen by international corporations, institutions, and governments.

During the Soviet Era, Russian and Uzbek were the languages used in higher educational

institutions; the Uzbek language and minority languages throughout the Turkestan region were used for secondary education. In 1992, Uzbek and the five other Central Asian languages (Kyrgyz, Tajik, Kazakh, Turkmen, Azerbaijani) were made the official languages of instruction at state schools, meaning that Uzbek schools might use any of five Central Asian languages or Russian as their primary language depending on their geographical location and the population of students. In other words, if a state school was close to Kazakhstan, the students would attend a Kazakh school, in which there are Uzbek nationals who have Kazakh heritage. There are Kazakh, Tajik, Kyrgyz, Turkmen, and Azerbaijani schools. At the same time, the five Central Asian countries (i.e., Azerbaijan, Kazakhstan, Kyrgyzstan, Turkmenistan, and Uzbekistan), agreed to adopt the Latin alphabet to improve relations amongst each other and make trade among the countries easier. With the addition of multiple languages and different writing systems, there arose much confusion among students and teachers. Thus, in 2005 the Cyrillic alphabet was officially abandoned and schools used the Romanized-Uzbek alphabet.

Also at the turn of independence, English instruction in schools was adopted – in addition to East Asian languages (i.e., Chinese) – and students could choose to study it in higher education institutions across Uzbekistan. This language placed new stress on a limited supply of teachers and materials, and language teachers who were originally teachers of Russian were asked (and some forced) to teach English. These teachers were expected to travel to Tashkent and take retraining courses. These courses were meant to support the English language development of teachers and introduce new language teaching methodologies, such as task-based language teaching, which was much different than the grammar-translation method used to teach the Russian language at Soviet-style institutions of higher education. The Vice Director of the National Center for Teaching Foreign Languages explained that the English language became

more common for students to study, and many policy makers expressed the hope that English would replace Russian as the language of international communication in Uzbekistan (K. Murad, personal communication, July 16, 2017). The infrastructure to support the development of English has been steadily growing since independence.

3.2.1 A Historic Moment

The U.S. Department of State considers Uzbekistan to be an authoritarian state because there was supposed to be a separation of powers among the executive, legislative, and judicial branches even though there was not a separation. Karimov, however, dominated all branches of government until he passed away on September 2, 2016. In the election that preceded his passing, former prime minister Mirziyoyev won with 88% of the vote. Many organizations, including the Organization for Security and Cooperation's Office for Democratic Institutions and Human Rights (OSCE/ODHIR), noted that "limits on fundamental freedoms undermine political pluralism and led to a campaign devoid of genuine competition (retrieved from <https://www.state.gov/j/drl/rls/hrrpt/2016/sca/265554.htm> on February 23rd, 2018)" However, the OSCE/ODHIR's report also identified positive changes such as "the election's increased transparency, service to disabled voters, and unfettered access for 600 international observers (retrieved from <https://www.state.gov/j/drl/rls/hrrpt/2016/sca/265554.htm> on February 23rd, 2018)" Since the election, the country has seen great changes and, from my own observations, people on the street are impressed with the government's transparency and recent decision making. On October 25, 2017, *The New York Times Magazine* published an article entitled, "Once closed and repressive, Uzbekistan is opening up." In this article, author Kramer discussed how President Mirziyoyev is spearheading a political and economic overhaul called the Uzbek Spring, which is meant to open up Uzbekistan to the world. This move of Mirziyoyev is also

seen in the December 2017 *Conde Nast Traveler* magazine with a fourteen-page spread of Uzbekistan promoting Uzbekistan as a travel destination in 2018. Not only is Uzbekistan opening up to the outside world, but its relationship with the United States is getting stronger.

The relationship between the United States and Uzbekistan has been tempestuous since 2005 when the U.S. reacted critically to the massacre of 1500 Uzbek nationals by the Uzbek government. The U.S. viewed the relationship with Uzbekistan as important because it was the country with the largest population in Central Asia and as the country that provided stability and security to the region; however, when this stability collapsed, the U.S. questioned whether they should continue bilateral relations. A Public Affairs Officer from the U.S. Department of State explained that the U.S. State Department decreased its humanitarian and technical assistance after the massacre, withdrew the U.S. Peace Corps because of the questionable safety of U.S. citizens, and argued “in favor of ending all U.S. ties to Uzbekistan” (N. Boltaeva, personal communication, July 2, 2017). At the same time, President Karimov then removed the military air base that was used by American troops to deploy to Afghanistan during the Iraq War. The political relationship between the two nations was tenuous, reflected in the 2012 U.S. Global Leadership Report findings that 40% of Uzbeks approved of the U.S. leadership, with 22% disapproving, and 39% uncertain.

The political relationship with the United States became more favorable after the death of President Karimov with the transition to President Mirziyoyev. This positive attitude toward the U.S. was not a result of the current U.S. administration but instead because of the change in the Uzbek regime. In an address to invited guests from the Uzbekistan government and selected U.S. citizens on July 4, 2017, U.S. Ambassador Pamela Spratlen talked about the relationship between the U.S. and Uzbekistan. In a speech, she explained that,

In a time of transition, it is also important to recognize the many things that remain constant, as we have done for more than 25 years. The United States continues to support the entitlements, sovereignty, and territorial integrity of Uzbekistan. I would like to review the major events of our bilateral relations this year... *our programs that support English language development continue to expand and develop* [emphasis mine] ...Dear guests, we are living in a historic moment in this country's beautiful history. The United States has been a key partner as Uzbekistan has begun its reform process. The government has signaled a desire for greater cooperation with the United States including through a robust road map. In the year ahead this and other initiatives, our Embassy and Washington colleagues stand ready to strengthen mutually beneficial bilateral and regional cooperation across all spheres (Chiesa, audio recording, 2017).

With Uzbekistan opening up to the world and with improving relations between the United States and Uzbekistan, I have had privileged access to this country while serving as an English Language Specialist for the U.S. Department of State. I have always looked forward to counseling and assisting U.S. mission officials in determining and developing the most effective use of English language programs and services to meet posts' Public Diplomacy and Mission goals. Thus, as an EL Specialist I had access to a group of underrepresented language teachers and decided to work with them to better assist the country of Uzbekistan in developing effective means for English language teaching and learning.

3.3 Establishment of a National In-Service Language Teaching Program

On June 15, 2015, Uzbekistan's President Karimov signed decree No. 24, "About measures for further improvement of the system of retraining and advanced training of managerial and pedagogical personnel in higher educational institutions" (see Appendix A for the English

version). This decree established an in-service training that lasted approximately three months and contained six modules divided into three overarching themes: (1) presidential decrees of Islam Karimov, (2) procedures at higher educational institutions, and (3) subject specific training. More specifically, the first five modules included compulsory topics on presidential decrees, education, law, as well as information technology in education, and management. Module six consisted of three main blocks and was tied to the teachers' special discipline. The curriculum for module six was determined by staff at each training site and specialists from within Uzbekistan and/or abroad.

Every three months, a select number of managerial and pedagogical personnel leave their position at their institutions to take part in the in-service training. According to one minister of education in Uzbekistan, the number of people to attend each session is determined by the physical layout of the training center and the annual budget (U. Azizov, personal communication, June 15, 2017). The Uzbekistan government financially supported each faculty and staff member in the travel costs, room, and board. The goal of the decree was to retrain all managerial and pedagogical personnel in higher education. The exact number of managerial and pedagogical personnel that will ultimately need to complete the process is not known.

Since the decree was signed in 2015, approximately 1,200 university language teachers from all twelve provinces, one autonomous region, and an independent city of Uzbekistan have come to the capital, Tashkent, to the Flying High Training Site. Flying High is a pseudonym for the training site that trains in-service language teachers. The pseudonym – Flying High – was given to protect the anonymity of all the stakeholders involved in training courses, including the trainers, administration, government agencies and NGOs who support the syllabus and thus, curriculum. The Flying High Training Site was assigned by the President of Uzbekistan as the

training site for language teachers of French, Spanish, German, Russian, and English. The retraining of teachers of Asian languages such as Chinese, Japanese, Korean, and Mongolian were retrained at a different site in Tashkent. At Flying High, most language teachers taught English as a Foreign Language (EFL). This research took place at the Flying High Training Site in 2017 with English university language teachers who attended one of the three-month in-service teacher education programs.

3.3.1 U.S. Embassy Role

The United States Embassy in Tashkent has supported programs of English language development since Uzbekistan declared English to be taught at the primary, secondary, and tertiary levels. A high-ranked Embassy employee explained that English language programs tend to be accepted by the Uzbek government more easily than other programs promoted by the Embassy (e.g., programs by the Public Affairs Section on American culture, such as Halloween). The U.S. Embassy does have the ability to reach the Uzbek public through presentations at the Information Resource Center and through Education USA, which are two departments within the Public Affairs Section of the U.S. Embassy. Ultimately, the day-to-day interactions of United States citizens and Uzbek nationals happens in language classes, with language teacher educators, and program administrators at colleges and universities. Foreigners are currently not allowed to work at primary and secondary levels in Uzbekistan.

I was assigned to work as an English Language Specialist for the United States Department of State in Uzbekistan at the Flying High Training Center in Tashkent for the mandatory in-service language teacher education program. I was assigned to revise the curricula of the in-service course, develop course materials, and conduct a series of training of trainers with the English language teachers. The sixth module, which consists of Teaching English as a

Foreign Language (TEFL), includes three blocks on (1) linguistics and second language teaching, (2) English teaching methodologies, and (3) language assessment. The focus of this dissertation is on the third block – language assessment. The dissertation project was separate from the EL specialist position; it was the language teachers who were enrolled in the program at Flying High Training Site that I was given permission to conduct research with.

3.3.2 Language Teachers at Flying High Training Site

Each teacher participated in the training for various reasons. Some teachers were told by their administration that they must go to Tashkent; some went of their own accord; a few teachers were contacted by the Flying High Training Site administration and told that they must attend because their higher educational institution had not sent any (or only a few) teachers since the implementation of the Presidential Decree in 2015. The Flying High Training Site kept records of which teachers attended and which teachers had not. Priority, however, was given to language teachers from one of the twenty-nine state-run higher educational institutions, and second priority was given to private institutions (A. Navoi, personal communication, July 19, 2017).

There was no systematic organization in place that determined which teachers were obligated at which times to attend the in-service teacher education program. All state universities had representation at the in-service training of this study. (See Figure 10 for a chart of the organizational structure of the different state-run colleges / universities, which is under the Ministry of Higher and Secondary Special Education of the Republic of Uzbekistan.)

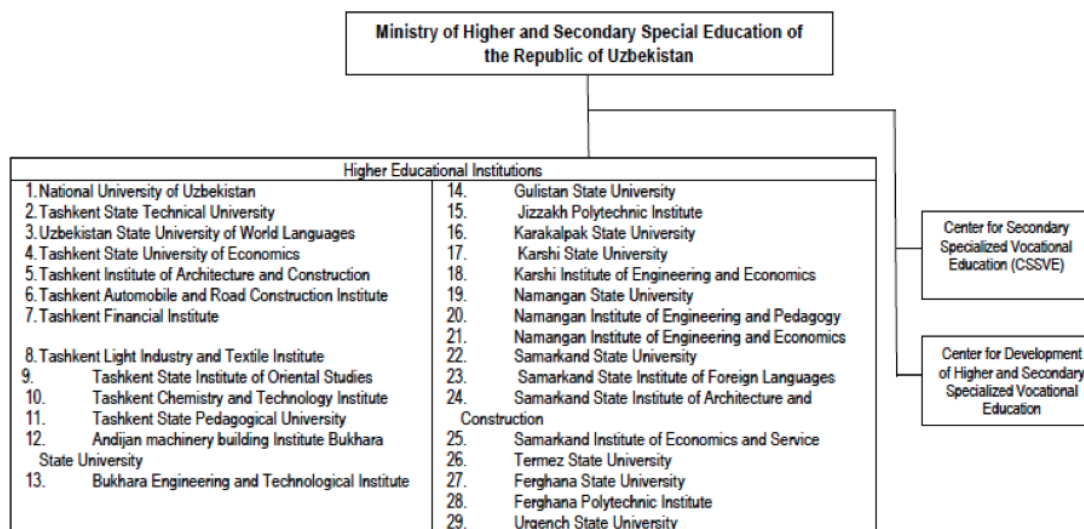


Figure 10 Organizational Structure of Higher Education Institutions

Figure 11 is a map of Uzbekistan, providing a graphical representation of where the capital is situated in relation to other major cities, towns, and villages. The capital, Tashkent, can be seen in the east of the country.



Figure 11 Map of Uzbekistan

Language teachers would travel several days by train to go to Tashkent if they worked in

Karakalpakistan in the northwest which is on the border of Uzbekistan and Kazakhstan, or if they lived in Termez in southern Uzbekistan bordering Afghanistan. The teachers came from all over the country and are thus not just a representation of Uzbekistan's capital city region.

The role the Flying High Training Site had in the recruitment of participants is considered an *engaged role* of a non-U.S. citizen organization, which is “one whose employees or agents participate in the recruitment of subjects, conduct the consent process, obtain consent from subjects, intervene with human subjects for research purposes, or receive the private, identifiable information of subjects” (<https://www.hhs.gov/ohrp/international/index.html>, retrieved on January 05, 2018). For me to be allowed to conduct this research, I had to agree to the recruitment process of the Flying High administration.

3.4 Engagement – Recruitment of participants

The country of Uzbekistan does not have its own IRB or comparable review committee, but relevant regulations are found in articles 24, 26, 40 and 44 of the Constitution of the Republic of Uzbekistan (1992). These articles establish the right of the Uzbek government to protect its citizens' rights from outside/foreign interference, influence, or research. U.S. Federal regulations do not require on-site, local review in such cases. However, the researcher and the IRB that reviews the research should be mindful of foreign regulations or other requirements that govern research on the local population.

One of the main contributions the Flying High Training Site administration was responsible for in the research process was the recruitment of participants. According to the director of the Flying High Training Site, all university English teachers were asked to complete the Language Assessment Literacy survey used to answer the first research question; however, the director understood that the teachers would need to sign a consent form to have their data be

used for research. For the second research question, which utilized focus group interviews, the director noted that he selected participants based on the region of Uzbekistan they came from, as he wanted to provide me with a representative sample of teachers who came from across the country. The Flying High Training Site administration did not assist in the recruitment of participants for research question three, which used the semi-structured interview protocol. These teachers came from the focus groups and volunteered their time to talk to me one-one-one.

3.5 Conclusion

In this chapter, I discussed the sociocultural background to the setting of the study. First, I explained that I chose Uzbekistan because it has had a rich history and culture that spans thousands of years and has had an impact on western countries in terms of religion, politics, and economics. Then, I showed how Uzbekistan's history, culture, and politics have influenced the development of English language teaching, particularly with Karimov's presidential decrees, that established the Flying High Training Site and the in-service teacher education program in which study participants were enrolled. Moreover, I discussed Uzbekistan's relationship with the United States, which has been a key factor in the development and implementation of the study. Finally, I explained how the government of Uzbekistan has had an engaged role in the implementation of the study, with the recruitment of the language teachers.

4 METHODS

There are three institutions involved in the process to protect the rights of the participating Uzbekistan EFL teachers, and all parties approved my research methodology before data were collected: Georgia State University, the U.S. Embassy in Tashkent, and the Uzbekistan Ministry of Foreign Affairs / Flying High Training Center. I received permission (see Appendix B) needed to conduct research and did not collect data beyond what I had permission to collect.

In this chapter, I discuss the research methodology used to address the following three overarching research questions and their subquestions, as I received permission for:

- 1) To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy?
 - A. What assessment skills and knowledge do Uzbekistan EFL teachers believe teachers need to possess?
 - B. Is the theoretical basis for the survey (Taylor's 2013 framework of LAL) faithfully implemented in the survey?
 - C. What is the factor structure present in the Uzbekistan EFL teachers' responses?
- 2) How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students?
 - A. What do Uzbekistan EFL university teachers report they do to assess their students (i.e., assessment tasks and scoring procedures)?
 - B. What are their cognitions surrounding those assessment practices?
 - C. How do their cognitions about assessment shape how they assign scores to their students' work?

- 3) What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students?
- A. What are the macro-environmental factors Uzbekistan EFL university teachers report?
 - B. What are their cognitions surrounding these factors?
 - C. How do the reported factors and cognitions shape Uzbekistan EFL university teachers' assessment practices?

I begin with an explication of the research participants. Then, I discuss the materials, data collection procedures, and the quantitative and qualitative analyses of research question one and its subquestions. Subsequently, I discuss research questions two and three and their subquestions with an explanation of the recruitment process, who the participants were, the interview protocols, and then, the qualitative analyses.

4.1 Research Participants

Ninety-nine university English language teachers from thirteen provinces participated in the study. Eighty-six teachers were women and thirteen teachers were men. Of these teachers, 56% taught at higher educational institutions in Tashkent, the capital of Uzbekistan, while 44% worked in the regions (i.e., Andijan, Bukhara, Fergana, Jizzakh, Xorazm, Namangan, Navoiy, Qashqadaryo, Samarkand, Sirdaryo, Surxondaryo, Tashkent region) and the autonomous region (i.e., Karakalpakistan).

Demographic information about the participants was collected when the Language Assessment Literacy Survey was distributed. Descriptive statistics of the participants' age and years' experience can be seen in Table 5. The median age of the participants was thirty-seven

with a mode of thirty-nine years old, and their ages ranged from twenty-six to sixty-three years old. The average age of the participants was thirty-seven. The median number of years of teaching experience at the university level was fourteen years with a mode of fifteen, and the number of years of experience ranged from one-year to forty years. The average number of years of teaching experience was eleven years.

Table 5 Descriptive Statistics of Teachers' Age and Teaching Experience (in years)

	<u>Mean</u>	<u>Median</u>	<u>Mode</u>	<u>Standard Deviation</u>	<u>Min</u>	<u>Max</u>	<u>Range</u>
Age	37	37	39	7.130	26	63	37
Experience	11	14	15	8.345	1	40	39

4.2 Research Question One

I administered Kremmel and Harding's Language Assessment Literacy Survey in order to gauge teachers' language assessment literacy. Additionally, I wanted to address the broader research question of the extent to which this particular survey, which purports to be based on a theoretical framework on LAL, actually provides valid and actionable information about teachers' LAL in a specific context. Specifically, I examined which assessment knowledge Uzbekistan EFL teachers believe language teachers (in general) need to know, and what assessment skills they think language teachers should possess. In addition to understanding the participants' cognitions about assessment, I also examined how valid the Language Assessment Literacy survey is for the Uzbekistan EFL teachers' language assessment literacy.

4.2.1 Data Collection Material – Adaptation of LAL Survey

I adapted Kremmel and Harding's (*forthcoming*) survey by reducing the number of items from 72 to 66 in order to include only the items that pertained to language teachers in an Uzbekistan

EFL context. (See Appendix C for the adapted survey.) The following items were not included in the survey I distributed²,

Table 6 Items Removed from Original Survey

Item Description
<i>how to use assessments to evaluate language programs</i>
<i>how to determine if a language assessment aligns with a local system of accreditation</i>
<i>how to train others about language assessment</i>
<i>how assessments can be used to enforce social policies (e.g., immigration)</i>
<i>the history of language assessment</i>
<i>accommodating candidates with disabilities or other learning impairments</i>

The items *how to use assessments to evaluate language programs* and *how to train others about language assessment* were removed because they were deemed irrelevant. In Uzbekistan, the people responsible for evaluating language programs and training others are expected to be specialists who work in the Ministry of Higher and Secondary Education. Additionally, three more items – *how to determine if a language assessment aligns with a local system of accreditation*; *how assessments can be used to enforce social policies (e.g., immigration)*; and, *accommodating candidates with disabilities or other learning impairments* – were removed because there are words in each of these items that cause concern: accreditation, immigration, and disabilities. Universities do not go through an accreditation process in Uzbekistan. The topic of immigration is considered a sensitive topic in Uzbekistan. The Flying High Training Site asked me to remove this item before distribution. Additionally, learners with *disabilities* and/or *learning impairments* are provided university-level education at a private institution. The language teachers who teach at this school are not asked to attend the in-service teacher education program at the Flying High Training Site. Finally, a sixth item – *the history of*

² Having eliminated these items, I cannot empirically demonstrate that these six items are not important to teachers.

language assessment – was removed for two reasons. First, this item received the lowest rating from Kremmel and Harding’s (2016) pilot study. Third, Harding reported he would remove this item for the final version of the survey (L. Harding, personal communication, October 16, 2016).

The adapted survey was composed of 66 short closed-item responses, which were divided into two parts. Part 1, items 1-44, measured how knowledgeable language teachers need to be about each aspect of language assessment. Teachers responded on a Likert scale from 0 (not knowledgeable at all) to 4 (extremely knowledgeable). Part II, items 45-66, measured how skilled language teachers need to be about each aspect of language assessment. Teachers responded on a Likert scale from 0 (not skilled at all) to 4 (extremely skilled). In addition to 66 closed items, a single open-response item and three demographic questions were included. The single open-ended item asked teachers whether specific items were written clearly. The biographical questions included gender, age, and years of teaching experience.

In addition to the English version, the survey was translated into Uzbek and Russian (see Appendix D for surveys and section 4.2.1.1 for the translation process) by qualified translators. These people were chosen because they were current members of the American Translators Association (ATA) and certified degree holders (MA) in translation and interpretation. Additionally, the translated Uzbek and Russian surveys went through an adjudication process. The translations of the survey from English to Uzbek and Russian, fulfilled the Georgia State University IRB requirement: to assist the university English teachers whose level of English is not proficient enough to complete the survey in English, and promote full completion rate of respondents.

4.2.1.1 Uzbek and Russian Translation of LAL Survey

Figure 12 depicts the process I undertook to translate the English version of the Language Assessment Literacy Survey into Uzbek and Russian.

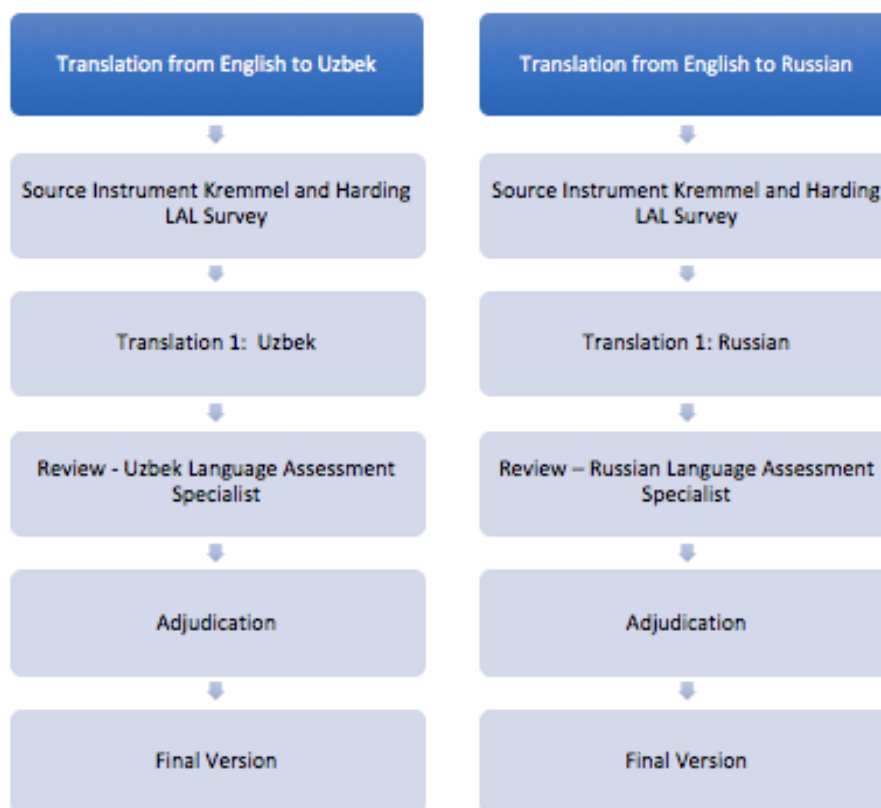


Figure 12 Translation of LAL Survey into Uzbek and Russian

First, I sent Kremmel and Harding’s Language Assessment Literacy Survey to a professional Uzbek translator. Subsequently, I sent the translated survey to an Uzbek language assessment specialist, who worked in the Language Assessment and Program Evaluation department at the Flying High Training Site. The director reviewed the translation and provided comments (see Table 7 below.) Subsequently, she had an email conversation with the professional translator, and a final version was agreed upon.

4.2.1.1.1 Ambiguity of Uzbek Translations

I sent the translated survey to an assessment specialist in Uzbekistan in May 2017, a month before I distributed the survey to the participating EFL teachers. She identified eight issues in the translation of the English version of the survey into Uzbek. Table 7 below presents the survey items, the Uzbek translation with the word or words bolded that may present a concern, and a written synopsis of the issues the assessment specialist identified.

Table 7 Uzbek Translation of Survey

Item	Uzbek Translation	Issue
8. <i>How to interpret measurement error</i>	Baholash xatosini qanday qilib sharhlash	The word “measurement” is translated as “assessment”.
17. <i>How to prepare learners to take language assessments</i>	Til bilimini baholashlarni o'tishga talabalarni qanday motivatsiya qilish	The word “how to prepare” is translated as “how to motivate”.
39. <i>Different types of purposes for language assessment (proficiency, achievement, diagnostic)</i>	Turli maqsadlarda tilni baholashning turlari (maslan, proficiency, achievement, diagnostic)	This item does not sound Uzbek. The word order must be changed. For example: <i>Tilni baholashning turlarining har hil maqsadlari.</i>
51. <i>Scoring closed-response questions (e.g., Multiple Choice Questions)</i>	...savollarni baholash (masalan, Multiple Choice Questions)	Here, “closed-response” was not translated at all.
52. <i>Scoring open-ended questions (e.g., short answer questions)</i>	...savollarni baholash (masalan, short-answer Questions)	Here, “open-ended” was not translated at all. It is written as “scoring question,” but no type of question is provided in Uzbek.
54. <i>Developing specifications (overall plans) for language assessments</i>	Til baholashlari uchun spesifikatsiyalarni (barcha rejalar) yozish	“Overall” is not correctly translated. Must be “umumiy”, not “barcha”
62. <i>Identifying assessment bias</i>	Baholashda yonbosishni aniqlash	The word bias does not exist in Uzbek.
66. <i>Selecting appropriate ready-made assessments</i>	To'g'ri tayyor baholashlarni tanlab olish	Here “ready-made” is not correctly translated. Maybe it must be “tayyor bulgan” but not “tugri tayyor”

The assessment specialist and the original/professional translator of the survey (who is Uzbek, has a professional translation degree, and an MA in TESOL) communicated about specific issues in the survey. The eight issues exemplify that the mistakes happened because there are no such words in Uzbek. The translator reported that “people just never use some of these words – such as *bias*, just not part of their culture, nor of their conceptual understanding” (personal communication, May 10, 2017). Therefore, the translator had a difficult time because he did not know what word, phrase, or clause equivalents to use. Many terms are also difficult: open-ended question, closed-response question, and bias. These must have a descriptive translation because there are no directly equivalent terms in Uzbek. Additionally, in an email the translator wrote to the Uzbek assessment specialist at the research site, he explained that in translation there are certain rules:

First rule, we have to stick to the original format as much as we can. The audience can have all kinds of background (well-educated or just-educated or not-educated). Long sentences are more for Level 3+ people. Sentences have to be simple here. Second, this is somebody's work. That person put thoughts, research and time to put this survey together.

It is not for us to decide to change the length and formatting (e-mail, May 25, 2017).

The relationship between the professional translator and the assessment specialist added another level of complexity, potentially affecting alignment with Taylor's (2013) original understanding. The differences in understanding from both a content knowledge specialist and a language specialist had to be negotiated to get the best possible translation to the original intended meaning.

4.2.1.1.2 Implementing the Uzbek Version of the Language Assessment Literacy Survey

A problem encountered in trying to use the Uzbek survey version was in its implementation. As a researcher in Uzbekistan, I had to be cognizant of the intricacies of the ethnic and linguistic background of its EFL teachers. I had prepared and received a professionally Uzbek translated version of the Language Assessment Literacy Survey, but did not have a Russian version. I assumed that with the departure of Russia in 1991, all EFL university teachers would be able to read and write in Uzbek and English. When I approached the first in-service EFL teacher class at the Flying High Training Site, which had been identified by the Uzbek Ministry of Education as the Euro class, I was unaware that the survey versions I had would not be helpful for the teachers. These EFL teachers did not read or write in Uzbek and they preferred to have had a Russian version. For the week following that situation, I had to work with a U.S. Department of State professional translator in Tashkent on an appropriate Russian version to be distributed, which then had to have approval from the Flying High Training Site and the Institutional Review Board at my home research base. For future use of surveys in Uzbekistan, it is prudent to provide all participants with multiple language versions, and then have the participants choose which one to use.

4.2.1.2 Russian Version

The translation of the Russian version of the survey did not happen simultaneously with the Uzbek translation. However, the Russian version “was easier to work with than the uncommonly used Uzbek language” (Nazutdinova [pseudonym], personal communication, May 2017). I present the findings of the ambiguous items below.

4.2.1.2.1 Ambiguity of Russian Translation

The word *local* and its collocations (e.g., local area, local customs) appeared in four survey items: Items 10, 12, 34, and 35 (see Table 8).

Table 8 Local Practices Items

Item	Item Description
10	<i>how to determine if a local assessment aligns with a local education system</i>
12	<i>how to determine if the results of the assessment are relevant to the local context</i>
34	<i>the assessment traditions in a local context</i>
35	<i>the relevant legal regulations for assessment in the local area</i>

These items presented some issues for the English-Russian translator. The direct translation of *local* in Russian is местный. However, for local practices, Russians may use the phrase местные обычаи / местные традиции (local traditions), or местные привычки (local habits). For local context, Russians may say местный контекст, but again, the Russian translator explained that “I think I never heard people say местный контекст in Russian. They would just say здесь, which means *here*” (personal communication, May 25, 2017). The number of options the English-Russian translator had to translate the word *local* is numerous. The denotations for each of the translated words are similar, but the connotations varied greatly because

these are really English-language word combinations, which we almost never really use in Uzbek or Russian languages. It happens so, I think, because English-language people work all over the world and it is important for them to identify local practices and local contexts in this or that place of the world. Because Uzbeks/Russians do not really work all over the world so much as English-speaking people do, these word combinations did not evaluate as fixed expressions as in the English language (Nazumutdinova, personal communication, May 25, 2018).

Therefore, if the participants chose the Russian version of the survey ($N = 29$), then they may have had difficulty identifying the intended meaning of *local*. With translation concerns in the survey translations from English to Uzbek and English to Russian, there arises a question of the survey's content validity and the overall design of the survey.

4.2.2 Data Collection Procedure

I distributed the survey to three separate groups at the Flying High Training Site. The first group were identified by the Ministry of Secondary and Higher Education as the "Euro" section ($N = 24$). These teachers were university-based language teachers who worked at Russian-focused (e.g., international) higher educational institutions throughout Uzbekistan. The survey was given to them on June 1, 2017.

The Euro group utilized the Russian and English versions of the survey, instead of the Uzbek version, because they were unable to read the Uzbek survey (personal communication, Euro group, June 1, 2017).

The second group self-identified as teachers who resided in Tashkent and the Republic of Karakalpakistan ($N = 43$), which is bordered by Kazakhstan in the north, Kyrgyzstan and Tajikistan in the east, Turkmenistan in the west and Afghanistan in the south. The survey was given to them on June 15, 2017. These teachers used all language variations of the survey, depending on which autonomous region they came from.

The third group of language teachers self-identified from the remaining eleven outside provinces (e.g., Andijan, Bukhara, Djizzak, Fergana, Kashkadarya, Khorezm, Namangan, Navoi, Samarkand, Surkhandarya, and Syrdara, $N = 32$) in Uzbekistan. The survey was given to them on June 29, 2017. These teachers used either the English or Uzbek versions of the survey, because they were unable to read the Russian document.

Each participant was provided three different versions of the survey – one in English, Russian, and Uzbek. The language teachers were instructed to select the version they felt most comfortable answering. Each teacher submitted only one survey for analysis. Of the ninety-nine participants, ninety-six submitted a version for analysis and interpretation, English ($N=45$), Russian ($N=29$), and Uzbek ($N=22$), which was a 97% return rate. As each group above were taking the Language Assessment Literacy Survey, I noted down any questions the participants had about a specific item, and which language version they were asking questions about. I also wrote down comments they made directly to me before and after they took the survey.

After each collection period, the data were then directly entered into a private and secure spreadsheet file. Each row represented a different survey participant (e.g., #1) while the columns represented each item in the survey (e.g., item1). The total number of items in the survey was sixty-six, with three additional biographical questions that asked about the teacher's gender, age, and years' experience in teaching English as a Foreign Language.

4.2.3 Analyzing the Language Assessment Literacy Survey

To analyze the first research subquestion – What assessment skills and knowledge do Uzbekistan EFL teachers believe teachers need to possess – I used the EFL teachers' answers to each closed-response item and Kremmel and Harding's factor analysis results. I used JASP v.0.8.3.1 – a free and open-source graphical program for statistical analyses – to compute the descriptive statistics for each item. I found the measures of central tendency (the mean, mode, and median) and measures of dispersion about the mean [range (minimum and maximum), standard deviation, and variance]. Graphical representations of histograms were created for salient items that the respondents endorsed strongly and those they did not endorse so strongly.

Then, I calculated descriptive statistics of teachers' responses, grouping items based on the nine factors Kremmel and Harding (*forthcoming*) extracted for their factor analysis ($N = >1000$). (See Table 9, which is ordered from factor one to nine.)

Table 9 Factors based on Kremmel and Harding (forthcoming)

Factor	Factor Name
One	Developing and Administering Language Assessments
Two	Assessment in Language Pedagogy
Three	Assessment Policy and Local Practice
Four	Personal Beliefs and Attitudes
Five	Statistical / Research Methods
Six	Assessment Principles and Interpretation
Seven	Language Structure, Use, and Development
Eight	Washback and Preparation
Nine	Scoring and Rating

I used Kremmel and Harding's results to gather evidence that can be used to evaluate claims of validity of the survey.

To answer the second research subquestion – Is the theoretical basis for the survey (Taylor's 2013 framework of LAL) faithfully implemented in the survey – I conducted an external review of the journal issue *Language Testing*, 30(3). I read closely and took notes of each article in *Language Testing* 30(3) to identify possible definitions that Taylor (2013) used to create her Language Assessment Literacy Framework, and which Harding and Kremmel used to write their survey items. Based on this analysis, I created my own operational definitions of each language assessment literacy dimension (Taylor, 2013): (1) knowledge of theory; (2) technical skills; (3) principles and concepts; (4) language pedagogy; (5) sociocultural values; (6) local practices; (7) personal beliefs/attitudes; and (8) scores and decision-making. Using my own definitions, I coded each Language Assessment Literacy survey item with two other coders. Then, I calculated inter-coder agreement and subsequently facilitated a focus group discussion about items of the survey that coders found to be ambiguous or too difficult to code. Follow-up

questions to each coder were addressed via email correspondence. I then determined a final coding scheme.

After I received Kremmel and Harding's official coding scheme, I compared my own with theirs, and discerned points of agreement and disagreement. These results provided information about the English version of the survey and whether it adhered faithfully to Taylor's (2013) framework. Because I used multiple language versions (e.g., English, Uzbek, and Russian), I then conducted an analysis to determine the potential translation issues that could have played a role across different language versions of the survey. Before I distributed the survey to the participants, I asked a language assessment expert in Uzbekistan to review the Uzbek version of the translated survey. She identified issues with the survey and then she and the professional translator came to an agreement about a finalized Uzbek version of the survey. The same process occurred with the Russian version of the Language Assessment Literacy survey.

To answer the third research subquestion – What is the factor structure present in the Uzbekistan EFL teachers' responses – I conducted an exploratory factor analysis (EFA) of the survey item response data and achieved a factor structure that was identified as having good fit indices and strong reliability. I then compared the factor structure extracted from my analysis with that of Kremmel and Harding's factor analysis results.

Finally, in order to address the overarching research question: To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy – I examined the results from all three subquestions above. This analysis determined if, indeed, the survey is a valid data collection instrument to be used for determining Uzbekistan EFL teachers' cognitions about language assessment literacy.

4.3 Research Questions Two and Three

Research questions two and three are meant to begin the process to unravel the complexity of the participating Uzbekistan EFL teachers' inner lives. Research question two (and its subquestions) identifies their reported cognitions at the micro-institutional level (e.g., classroom and university context). I ask the following: What do Uzbekistan EFL university teachers report they do to assess their students (i.e., assessment tasks and scoring procedures); what are their cognitions surrounding those assessment practices; and, how do their cognitions about assessment shape how they assign scores to their students' work? Similar to research question two, research question three (and subquestions) identifies their reported cognitions. Instead, the focus of the question is at the macro-environmental level (e.g., sociocultural and sociopolitical). I ask the following: What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students; what are the macro-environmental factors Uzbekistan EFL university teachers report; what are their cognitions surrounding these factors; and, how do the reported factors and cognitions shape Uzbekistan EFL university teachers' assessment practices. To answer these questions, I conducted focus-group and semi-structured interviews.

4.3.1 Focus group interviews.

The focus group interview was selected because it allowed me to gain a variety of perspectives and experiences from the participating Uzbekistan EFL teachers across the country. Secondly, the focus group acted as "a compromise between participant observation and individual interviewing as it is less controlled than traditional interviewing and more controlled than participant observation" (Ho, 2012, p. 2). Additionally, I began with the use of focus groups instead of semi-structured interviews to establish trust with the teachers who had seen me around

the Flying High Training Site. Finally, the focus group research method was chosen because of my two prior experiences conducting focus groups for a professional development organization (American Association for Applied Linguistics) and with prior applied linguistics research.

4.3.1.1 Focus group recruitment and participants.

The teachers were selected by the Flying High Training Site administration. The administration also made a timetable schedule, scheduled a room, and then notified me of that information. No members of the Flying High administration were present during the focus group interviews.

There were five different focus group interviews, each with ten or eleven participants; one male teacher was in each group with the other participants all female teachers. The participants came from each province and autonomous region in Uzbekistan. Most language teachers were from Tashkent ($N=19$), then Karakalpakistan ($N=8$), Fergana ($N=7$), Namangan ($N=4$), Samarkand ($N=3$), Andijan ($N=2$), Bukhara ($N=2$), Djizzak ($N=2$), Khorezm ($N=2$), Kashkadarya ($N=1$), Navoi ($N=1$), Surkhandarya ($N=1$), and Syrdara ($N=1$). I provided each interviewee with an Uzbek pseudonym to protect their identity. (See Appendix E for Uzbek Pseudonyms.)

4.3.1.2 Focus group materials

The focus group interview consisted of ten questions: (1) Do you have any kind of set curriculum or can you decide for yourself what you will teach and assess? (2) What are your learning goals for your students, and what are your students' goals / Do these goals ever conflict? (3) In your classes, how are the grades determined for each student? (4) Have you had any training in testing/assessment? (5) At each of your universities/colleges, how do you track students' (spoken, written, listening, reading, grammar, etc.) language learning? (6) How do you decide what to test? (7) How do you make tests? (8) What do you do with the results of the tests? (9) If

a student does not pass a test or class, what are some consequences? (10) Describe a challenge you have had in assessing your students. (See Appendix F for full focus group interview protocol.) The focus group data was collected during the last week of June and first week of July 2017.

4.3.1.3 Focus group procedure

I arrived early before each session and made sure the room was set up with one round table and enough chairs. Then I made sure that the consent forms were printed out and that there were enough pens for each teacher to sign his/her name and extra paper for name cards. Finally, when the language teachers arrived, I greeted them personally to create an open and friendly atmosphere.

At the start of the focus group, I read through the consent form and had each teacher sign and ask me questions. Part of the consent form was the explanation that I would audio-record the conversation and wanted to make sure all teachers felt comfortable with the recording. No teacher opposed being recorded. Then each teacher was asked to write the name of the province she/he worked in on a name card. If more than one teacher came from a province, then the teacher would put #1, #2. after the province name. I also instructed the teachers not to state their real names or the name of the school they were working in, to protect the teacher's anonymity with the readers of this research and those other participants at the focus group. After the initial greeting and instructions, I asked the participants the set of ten questions above including clarification and repetition questions. In addition to asking questions, I also took written notes as a coping strategy for keeping the focus group running smoothly. Finally, at the end of each focus group interview, I summarized what was asked and discussed during the session. The shortest

duration was 38 minutes and the longest was 48 minutes, with an average time of 43 minutes.

Table 10 shows the duration of each session.

Table 10 Focus Group Interview Time Table

<u>Focus Group Number</u>	<u>Length in Minutes and Seconds</u>
Focus Group 1	45:33
Focus Group 2	47:54
Focus Group 3	38:00
Focus Group 4	34:12
Focus Group 5	48:05

Over the next few days I listened to the recordings and transcribed verbatim what was said in the discussion. The transcription software Dragon v. 5.0.0. for Mac was used to assist in the transcription process. I would play the recording next to the microphone on the computer, then Dragon software transcribed loosely what was heard into a Microsoft Word document. After each section of the recording, I would relisten and then make necessary corrections that the computer software could not pick up or misinterpreted, which occasionally were the names of places (e.g., provinces). One of the difficulties of transcribing focus group interviews was the common occurrence of overlapping multiple voices of the participants and deciding who said what and when. Dragon software eased the process of transcription because it has a feature of voice detection, which make it easier to identify different participants when they spoke. Once all transcriptions were completed, they were placed in an encrypted file on a computer with the audio recordings.

4.3.2. Semi-structured interview

The semi-structured interview was conducted after the focus groups, because this method yielded complex data with an emphasis on subjectivity involving the Uzbek language teachers' stories, images, and descriptions. Part of this process was the sharing of sensitive information that they would not have wanted to share during the focus group interview format. Additionally, semi-

structured interview procedures were employed because they provided me with an avenue to ask clarifying questions or ask for a teacher to expand upon a point. The purpose was to uncover what the teachers do and why they do what they do for language assessment. The semi-structured interviews (and focus group interviews) provided answers to the “why” question, or in this study, the cognitions behind what they did with certain assessment practices.

4.3.2.1 Recruitment and participants. The recruitment for the semi-structured interview was based on who participated in the focus-group interviews. At the end of each focus group, I informed all participants of the third part of the study, which involved a one-on-one interview, and then gave each of them my student and personal email accounts. I asked them to contact me if they would like to talk further about assessment, and that the interviews were completely voluntary and confidential.

Of the fifty-three participants in the focus group, forty of those teachers contacted me by email and wanted to have a personal semi-structured interview. Because of the limited time and resources available to me to conduct all forty interviews, I selected participants based on region, and chose twelve teachers. Thus, I conducted twelve semi-structured interviews with one person from each province and the autonomous region of Karakalpakistan. Of the twelve participants, eleven were female and one was male. These teachers had been teaching English at the university level for approximately eight to fifteen years.

4.3.2.2 Semi-structured interview protocol and procedure. The semi-structured interview consisted of eight questions (see Table 11 below for questions and Appendix G for full Semi-Structured Interview Protocol):

Table 11 Semi-Structured Interview Questions

Can you please tell me about a time in your life / an experience when you took a high-stakes language test? How did you feel?
Can you tell me about an interesting experience when you were trying to assess a student's language abilities?
Can you remember how you felt when YOU were being assessed/tested?
Are you respected more by your peers if your students do well on a test? How is language testing looked at by your administration, at your university?
Do students get/receive more money or prestige if they do well on language tests? What happens if they do not do well?
What are the economic benefits to students of doing well in school?
How large a role does the department head, dean, or higher-up administrative person have in the creation, implementation, and scoring of your language tests?
Can you talk about/discuss any Uzbekistani cultural rules (spoken and not spoken) for assessing students?

All interviews were conducted in July 2017. The procedure consisted of three separate steps. I arrived at my office at the Flying High Training Site early before each scheduled interview, with the day and time agreed upon by me and each teacher. All interviews were in a private, quiet, individual office: my office at the training site. I made sure that the IRB consent form was printed out for each interviewee to sign and date. Finally, when each participant arrived, I greeted them and thanked them.

At the start of the interview, I read through the consent form and had each teacher sign the form and ask me any questions about the interview procedure. (For the teachers who participated in the semi-structured interview, they heard the purpose of the study two times before – once during the survey and again for the focus group interview. All teachers were aware of the purpose of the study and wanted to participate.) As in the focus group interviews, I explained to each teacher that I would audio-record the conversation and take written notes. No teacher opposed being recorded, and they all allowed me to take written notes in order to keep

the conversation going and to help track substantive matters. I also instructed each teacher not to state their real name or the name of the school they were working in, to protect their anonymity. After the initial greeting and instructions, I asked the participants the set of eight questions above including clarification and repetition questions. Finally, at the end of interview, I summarized what was asked and discussed during the session. The length of time for each interview varied. The shortest time was just over 12 minutes, while the longest was 1 hour and 8 minutes. The total amount of recorded time was 7 hours and 55 minutes, with an average interview time of 40 minutes.

I provided each interviewer with an Uzbek pseudonym to protect their identity. (See below for Table 12, that shows the province the teacher came from, the length of the interview, and the pseudonym given to the teacher.) These pseudonyms are the same as the ones from the focus groups. I used the same pseudonyms to track their reported cognitions from the focus group to the semi-structured interview for data analysis.

Table 12 Semi-Structured Interview Information

	<u>Pseudonym</u>	<u>Province of Work</u>	<u>Length of Interview</u>
Interview 1	Klara	Tashkent	44:59
Interview 2	Svetlana	Samarkand	31:13
Interview 3	Ulugbek	Djizzak	32:16
Interview 4	Diora	Khorezm	41:01
Interview 5	Umida	Karakalpakistan	34:07
Interview 6	Nodira	Andijan	1:05:36
Interview 7	Aziza	Fergana	32:10
Interview 8	Kamila	Navoi	32:07
Interview 9	Mohira	Kashkadarya	34:30
Interview 10	Feruza	Namangan	12:01
Interview 11	Nozliya	Syrdara	24:39
Interview 12	Mukaddas	Bukhara	1:08:15

Over the months of July, August, September, and October 2017, I listened to the recordings and transcribed verbatim what was said in each interview. As with the focus group interviews, the same transcription procedure with the software Dragon v. 5.0.0. for Mac was used.

4.3.2 Analysis of Focus Group and Semi-Structured Interviews

Research questions two and three yielded qualitative data, and both involved the same analyses. The analyses consisted of *open* or *substantive coding*, which is when researchers “read their data and begin to code all different types of actions, events, processes... and write analysis notes in the margins of the documents they are working on” (Dillon, 2013, p. 3). This type of coding is also referred to as content analysis. In content analysis, key points are identified from the data. “Content Analysis is an interpretive approach and therefore includes extracting examples from the data” (Ho, 2012, p. 2). Thus, two separate files were created in NVivo (v. 10.2.2) for Mac, to analyze questions two and three and their subquestions.

After combing through the focus group and semi-structured transcripts, each line(s) or paragraph was coded for research question two: What do Uzbekistan EFL university teachers report they do to assess their students (i.e., assessment tasks and scoring procedures)? This question began the qualitative data analysis process, which started with coding for how teachers reported how they do *Assessment-of-Learning* and *Assessment-for-learning* practices. Then, subcategories were created from the broad categories after the data had been read multiple times. Subsequently, to address the inquiry – What are their cognitions surrounding those assessment practices – I went back through the introduction phase of each focus group and semi-structured interview and charted each teachers’ self-introduction. I noted down the following: pseudonym of each teacher, what province s/he came from, what type of university s/he taught at, what type of curriculum s/he are a part of, and as much of their background as they revealed throughout the

interviews. From these identifiers, I reread the transcripts and matched who said what at what point to discern patterns in the teachers' cognitions and their reported assessment practices. To address the third subquestion – How do their cognitions about assessment shape how they assign scores to their students' work – I combed back through the transcripts and identified how each teacher (or a group of participating Uzbekistan EFL teachers) scored students. Based on the results of the three subquestions I determined how the teachers' use assessment for the service of student learning.

After I addressed research question two (and its subquestions) I analyzed research question three. I began with the following inquiry: What are the macro-environmental factors Uzbekistan EFL university teachers report? Lines and paragraphs in the focus group and semi-structured interview transcripts was coded for social, cultural, economic, and/or political factors that influenced the Uzbekistan EFL teachers' assessment practices. Following the initial coding of the data, I read through the data six more times for possible coding categories that could be merged, modified, or clarified, and then identified two core codes. To make sure the core codes were consistent, I utilized the card-sort technique, initially developed by Lincoln and Guba (1985) as a quality control mechanism. I placed individual statements from each category onto three-by-five index cards and placed them randomly in a pile. After I explained the definitions of each category to two additional coders (one for research question two and three respectively), I had them distribute all the cards into the respective categories. Then, *intercoder agreement* was calculated. Then, to answer the second subquestion - What are their cognitions surrounding these (macro-environmental) factors – I used the charted categories from research question two (see above) and matched which teacher said what about each macro-environmental factor and when. After identifying the patterns in the teachers' cognitions, I answered the overarching research

question that inquired about the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students.

4.4 Conclusion

In this chapter, I discussed the research methodology used to address three overarching research questions (and nine subquestions): (1) To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy; (2) How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students; and, (3) What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students? I began with an explication of the research participants, which included descriptive statistics of their ages and number of years of experience teaching English in university settings. Then, I discussed research question one, and how Harding and Kremmel's Language Assessment Literacy survey was adapted for this study, followed by the procedure for its distribution and quantitative analysis. Furthermore, I explained the necessary steps to validate the Language Assessment Literacy survey.

Subsequently, I discussed the methods of research questions two and three. First, I explained the recruitment process for the focus groups, the focus group interview protocol, and the procedure for carrying out the focus group interviews. Then, I discussed the semi-structured interviews, the recruitment process, and the interview protocol. Because research questions two and three yielded qualitative data, I explained how they were both analyzed similarly, which began with open (substantive) coding. The results from the three research questions will present a picture of

if and how the participating Uzbekistan EFL teachers provide assessment at the service of learning for students.

5 RESEARCHER POSITONALITY AND REFLEXIVITY

In this chapter I will address my positionality, which I believe is a valuable and necessary exercise for any researcher who collects and analyzes empirical data. One of the challenges of writing this statement is to be ethically honest and open, so I would be able to acknowledge biases that are inherent in my orientation toward research, the research process, the content under investigation, the participants, and the analyses. The purposes of creating a positionality statement cannot be overstated because it will inform me, the researcher, and you, the reader, that I am cognizant of my own biases and I am trying to address them so I can achieve something as close as possible to what is commonly referred to as ‘truth’ from the research process.

Smith (2013) explains that “research is not an innocent or distant exercise, but an activity that has something at stake, and it occurs in a set of political and social conditions” (p. 15). The conditions she referred to are all related to issues of power. Those of us who serve as researchers are exercising a form of power when we engage with participants, the questions that we write or ask, and particularly when we assign categories to data collected. To address my positionality and to analyze such power relationships, I will locate myself in relation to the phenomenon under investigation – the cognitions of language teachers related to assessment literacy. Then I will locate myself in relation to Uzbekistan EFL teachers, who are the study’s focal participants. Finally, I will locate myself in relation to the research context and process of conducting qualitative research. Throughout, I will address how I have tried to overcome the barriers of my position as a researcher in the field of Applied Linguistics.

5.1 Researcher position in relation to the phenomenon

In my previous role as an English Language Fellow, working for the U.S. Department of State in the East Asian Pacific (EAP) region, I assisted EFL teachers in developing their knowledge and

skill base of language assessment. Through our many discussions, I learned about these teachers' perceptions of testing and their beliefs concerning the influence that classroom-based and high-stakes exams have on students and society. I have tried my best to maintain cognizance that my collaboration with EAP EFL teachers influences my perceptions of Uzbekistan EFL teachers' cognitions concerning assessment practices. Throughout the data collection and analysis process, I consistently reminded myself I am working in a new context, with teachers from a different cultural background, who come to the language classroom from a different cultural perspective, and who have different outside influences. I believe that my awareness of such influences helped me be more sensitive to and aware of the participating EFL teachers' cognitions on language assessment. Most importantly, when I analyzed the data, I made concerted efforts to allow the data to speak for itself instead of me looking for answers or patterns based on my past experiences with EAP EFL teachers.

5.2 Researcher position in relation to the training site's stakeholders

The director of the Flying High Training Site introduced me to the participating Uzbekistan EFL teachers. He identified me as "the U.S. Department of State's English Language Specialist, who is a Ph.D. candidate in applied linguistics". There were different phrases in this introduction that positioned me in relation to the participants, which might have influenced the data I collected. Let me begin with the phrase "Ph.D. candidate." This phrase is important because the people of Uzbekistan place great value on educational status, and I was often addressed by them as "dear professor." I tried to position myself on more of an equal level with the participants, even though our academic positions were different. I emphasized I am not a professor and that this particular nomenclature and its associated status should not be placed on me. However, the participants continued to call me professor, and they often made comments about how much they value and

respect research. Given this recurring interpersonal experience, how the teachers perceived me and my status could have influenced the type of data I collected.

The second phrase that was very important to understand my positionality was the term “English Language Specialist.” The Director explained to the participating Uzbekistan EFL teachers I had been assigned to come to Uzbekistan with the U.S. Department of State in order to serve as a “change agent.” Specifically, I had come to make any relevant recommendations for change to the national in-service training of English language teachers (see Chapter 3). This positionality could have affected the type of data I collected in two ways. First, the participants could have provided me with responses that would push their own agendas on the development of the in-service teacher education program. Second, the participants’ responses could have a pro-U.S. perspective as they may have wanted me to perceive them favorably on both personal and professional dimensions. This phrase, *U.S. Department of State*, carries with it some rather overt political connotations, and some teachers were curious as to my exact relationship with the U.S. government. Through my professional behaviors and interactions with them, I tried to demonstrate that my identity as an English Language Specialist was different from that of a non-academic U.S. Embassy employee, even though one of my major duties was to work effectively with embassy and professional colleagues from host country institutions to design, implement, evaluate, and promote English language programs in a cross-cultural environment.

When I analyzed the data, I was aware of the positive spin some participating Uzbekistan EFL teachers featured in their responses to my questions. The participants did not speak negatively about the U.S. government or those affiliated with their specific programs. To identify the teachers’ cognitions about foreign influences in their country, I tried to make connections between the countries that were mentioned (e.g., U.S. and U.K.). Thus, when U.S. government or

EL Specialist is mentioned in the results, I am aware that my positionality/affiliation with the EL Specialist program and the U.S. Department of State may have affected the data I collected.

5.3 Researcher position in relation to the investigative context and process

I acknowledge that research will necessarily influence and be influenced by the research context. This phenomenon is commonly discussed in the research method literature as the reactivity effect. My status as an English Language Specialist was intertwined with my positionality as a researcher, and I took on both emic and etic perspectives during this process. An *emic* perspective reflects an insider's view, meaning that I "discover the culturally specific frameworks used by members of a society/culture for interpreting and assigning meaning to experiences" (Pike, 1954, p. 579). In other words, adopting an emic perspective is an attempt to capture the culture member's point of view. This perspective was facilitated through my day-to-day interactions with the language teachers and in the words they used to discuss their culture and society. "Emic analyses incorporate the participants' perspectives and interpretations in the descriptive language they themselves use" (Nunan & Bailey, 2009, p. 197). Adopting an *etic* perspective includes doing one's best to maintain an outsider's point of view. In other words, when researchers take an etic perspective they often use preexisting theories or hypotheses to see if they are applicable to other existing settings or cultures. For instance, I took an etic perspective to the study when I analyzed the data. I used preexisting theories on L2 teacher cognition and language assessment literacy (see Chapter 2, Literature review) as a guiding framework to see if they applied to the Uzbekistan EFL context. Lett (1990) defines etic constructs as "accounts, descriptions, and analyses expressed in terms of the conceptual schemes and categories regarded as meaningful and appropriate by the community of scientific observers" (p. 130).

During the first month, I purposely sought to operate within an emic perspective by doing my best to have candid conversations with informants, attend English teaching development seminars with teachers, and attend Uzbek cultural events (e.g., weddings). Additionally, while conducting my job as an EL Specialist, I was asked by the Flying High Training Site administration to observe all teacher training classes at the center and talk to each teacher individually. The administration wanted me to examine how teacher development occurred in Uzbekistan and to gain insights into the participants' thoughts about the teacher education program. My interactions with the participating EFL teachers extended to interactions in the hallways and my office and going out to lunch together. One major interaction I had with all language teachers was when I was asked by the Flying High administration to provide language assessment professional developmental workshops (e.g., Using Role Plays to Assess Speaking; Assessing Speaking; Conducting Inter-Rater Reliability). The above interactions I had with the Uzbekistan EFL teachers occurred before I started collecting data. Thus, I believe I was able to establish a requisite level of positive relationships with the participants and gain insights into how they thought about Uzbekistan's education system as well as their impressions about how the English language is taught, learned, and assessed in Uzbekistan.

The understanding of my research positioning was important as I analyzed the qualitative data from my focus group and semi-structured interviews. I was cognizant that what I was told (and observed) outside of the data collection process would not be a part of the study's final results. Thus, I realized I needed to interpret study findings based only upon the information that I could collect, and which the language teachers had agreed to when they signed-off on the consent form.

5.4 Conclusion of Positionality

This chapter explained my positionality as a researcher, which is in direct relationship to my position at the Flying High Training Site. Thus, it should be noted that the results section will have some bias, but, I have explained the steps taken to mitigate and/or lessen the impact of my biases on three different levels. First, I located myself in relation to the phenomenon under investigation – the cognitions of language teachers related to assessment literacy – and tried to be aware when I was analyzing the data not to be influenced by my previous experiences in working with Uzbek teachers on their assessment practices. Then, I located myself in relation to the participating Uzbekistan EFL teachers, who were the study’s focal participants. My status as an English Language Specialist for the U.S. Department of State and a Ph.D. candidate in applied linguistics, positioned me in a particular way within Uzbek culture (e.g., a level of relatively high prestige). Uzbeks respect academic statuses; on a daily basis I explained to the participants why I was in Uzbekistan, my reasons for coming to their country, how I was related to the U.S. Embassy (e.g., as an academic; not as a diplomat), and ultimately, my passion for the teaching and learning of languages. My status was important to acknowledge because it influenced what and how information was presented to me by the study participants during the data collection process, particularly within the focus-group and semi-structured interviews. Finally, I located myself in relation to the research context and process of conducting qualitative research. In an effort to help to solidify the results of the qualitative section, a Ph.D. colleague assisted by joining me in conducting inter-coder reliability checks on the data. This principled collaboration with a research colleague helped me to focus on themes revealed by data only, and not to be influenced by outside factors that I experienced (or heard from teachers that were not part of the data collection process). In this section, I trust I have demonstrated that I have considered my

position and understand the power differences I had with the participating Uzbekistan EFL teachers.

6 RESEARCH QUESTION ONE RESULTS AND DISCUSSION

In this chapter I report the results of the first research question, To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy? To answer this inquiry, I address the following three questions: A. What assessment skills and knowledge do Uzbekistan EFL teachers believe teachers need to possess? B. Is the theoretical basis for the survey (Taylor's 2013 framework of LAL) faithfully implemented in the survey? C. What is the factor structure present in the Uzbekistan EFL teachers' responses?

From the EFL teachers' responses, I calculated descriptive statistics for each item, the overall survey, and then computed descriptive statistics for grouped items based on the nine factors Kremmel and Harding (*forthcoming*) extracted for their factor analysis ($N = >1000$). These results generally explained Uzbekistan EFL teachers' cognitions about which assessment knowledge they believe language teachers (in general) need to know, and what assessment skills they think language teachers should possess. To answer the second question, I conducted an external review of the journal issue *Language Testing*, 30(3) – a special journal issues focusing on language assessment literacy – to identify possible definitions that Taylor (2013) used to create her Language Assessment Literacy Framework, and which Kremmel and Harding used to write their survey items. I then created my own operational definitions of each language assessment literacy dimension (Taylor, 2013) and coded each survey item with two other coders ($\alpha = 0.91$). After I received Kremmel and Harding's official coding scheme, I compared my categorization with theirs, and discerned points of agreement and disagreement. These results provided information about the English version of the survey and whether it adhered faithfully to Taylor's (2013) framework. To answer the third question, I conducted an Exploratory Factor

Analysis (EFA) to identify the factor structure inherent in the teachers' responses. Subsequently, I compared my results with Kremmel and Harding's EFA to determine if the participating Uzbekistan EFL teachers responded to the survey items similarly with other language assessment stakeholders around the world.

6.1 Overall Numerical Summary

Uzbekistan EFL teachers identified all items in the survey as important (or somewhat important) for all language teachers to be able to know and do. Table 13 presents the descriptive statistics of the full survey, and Appendix H shows each item's descriptive statistics. For each item, a 0 (not knowledgeable at all, not skilled at all) to 4 (extremely knowledgeable, extremely skilled) Likert scale was used. In general, items had a mean score of 2.78, with an average median and mode of 3.00, with a minimum item mean of 2.35 and maximum item mean of 3.23. The standard deviation of item means is 0.16. This result suggests that the teachers perceive all facets of language assessment knowledge and skills (and their attitude toward it) to be important or somewhat important to a similar degree.

Table 13 Descriptive Statistics

	Mean	Median	Mode	Standard Deviation	Min	Max	Range
AL Survey	2.78	3.00	3.00	0.16	2.35	3.23	0.88

6.1.1 Measures of Central Tendency

Table 14 is a summary of the measures of central tendency.

Table 14 Measures of Central Tendency

	Mean			Median		Mode	
	$\bar{x} < 2.50$	$2.50 < \bar{x} < 3.00$	$\bar{x} > 3.00$	2.00	3.00	2.00	3.00
# of Items	1	59	6	5	61	6	60

Note. 66 items in total.

Fifty-nine of the 66 items had mean scores between 2.50 and 3.00, with just one item falling below 2.50 and six items with an average score above 3.0. Sixty-one items had a median of 3.00, while five items had a median of 2.00. Sixty of the 66 items had a mode of 3.00, while six items had a mode of 2.00. I will discuss individual items with the highest and lowest means below.

6.1.2 Measures of Dispersion

Table 15 is a summary of the measures of dispersion.

Table 15 Measures of Dispersion

	SD		Range		Maximum	Skewness
	0.76 – 0.99	1.00 – 1.07	3.00	4.00	4.00	[-0.09] – [-1.04]
# of Items	56	10	20	46	66	66

Note. 66 items in total.

The standard deviations for each item ranged from 0.76 to 1.07, where ten of the 66 items had standard deviations above 1.00, and fifty-six items had a standard deviation between 0.76 and 0.99. Each item had a reported value of 4.00 for a maximum score. Twenty items had an average range of 3.00, while forty-six of the 66 items had an average range of 4.00. Each item had a negative skewness (ranging from -0.09 to -1.04), suggesting that the item responses were distributed toward the higher end of the Likert scale. The measures of dispersion, similar to the measures of central tendency, emphasize the lack of variability in the data set and suggest that participants as a group perceived the importance of the survey items similarly.

6.1.3 General Categories Numerical Summary

I calculated descriptive statistics of teachers' responses, grouping items based on the nine factors Kremmel and Harding (*forthcoming*) extracted for their factor analysis ($N > 1000$). (See Table 16, which is ordered from highest to lowest mean.)

Table 16 Descriptive Statistics – Factors based on Kremmel and Harding (forthcoming)

Factor Rank	Factor Name	k	Mean	SD
Eight	Washback and Preparation	4	2.98	0.12
Seven	Language Structure, Use, and Development	5	2.97	0.18
Two	Assessment in Language Pedagogy	5	2.91	0.08
Six	Assessment Principles and Interpretation	4	2.85	0.14
Nine	Scoring and Rating	3	2.80	0.02
One	Developing and Administering Language Assessments	12	2.79	0.09
Five	Statistical / Research Methods	4	2.62	0.14
Three	Assessment Policy and Local Practice	4	2.60	0.05
Four	Personal Beliefs and Attitudes	4	2.55	0.14

The three categories participants reported needing to know the most about were Washback and Preparation ($k = 4$, $M = 2.98$, $SD = 0.12$), Language Structure, Use, and Development ($k = 5$, $M = 2.97$, $SD = 0.18$), and Assessment in Language Pedagogy ($k = 5$, $M = 2.91$, $SD = 0.08$). The three factors that received attention in the middle of importance were Assessment Principles and Interpretation ($k = 4$, $M = 2.85$, $SD = 0.14$), Scoring and Rating ($k = 3$, $M = 2.80$, $SD = 0.02$), and Developing and Administering Language Assessments ($k = 12$, $M = 2.79$, $SD = 0.09$). The bottom three factors that received relatively lower attention, but were still seen as important, were Statistical / Research Methods ($k = 4$, $M = 2.62$, $SD = 0.14$), Assessment Policy and Local Practice ($k = 4$, $M = 2.60$, $SD = 0.05$), and Personal Beliefs and Attitudes ($k = 4$, $M = 2.55$, $SD = 0.14$). Generally, participants viewed each factor as important or somewhat important for language teachers to know and possess. There are salient items that are worth noting from the survey responses.

6.1.4 Salient Items (Strongly and Not Strongly Endorsed)

For ease of reading, the presentation of items will have three parts: first, the number of the item from the survey; second, the written item in italics; third, in parentheses the number of respondents, the mean, and the standard deviation. Here is an example of prose: “The descriptive

statistics show that, on average, Item 23 – *how language skills develop* ($N = 96, M = 3.23, SD = 0.76$) ...”

6.1.4.1 Items that were Strongly Endorsed

Participants strongly endorsed items that addressed language, the relationship between teaching and assessing, and providing feedback. The participants provided a mean score higher than 3.00 for Items 3, 19, 21, 22, 23, and 30 (see Table 17). Of these, four items (i.e., 3, 19, 21, 23) had the smallest standard deviations in the data set.

Table 17 Highest Means and Lowest Standard Deviations

Item	Item Description	Mean	SD
23	<i>how language skills develop</i>	3.23	0.76
22	<i>how assessments can influence teaching and learning in the classroom</i>	3.09	0.81
21	<i>how assessments can influence teaching and learning materials</i>	3.06	0.78
19	<i>how to give useful feedback on the basis of assessment</i>	3.05	0.80
3	<i>how to use assessments to evaluate achievement in language learning</i>	3.04	0.78
30	<i>the structure of language</i>	3.03	0.84

Based on Kremmel and Harding’s original categorization, Uzbekistan EFL teachers strongly endorsed two items that were coded for Knowledge of Theory and four items coded for Language Pedagogy. The two items coded under Knowledge of Theory are Item 23 – *how language skills develop* ($N = 96, M = 3.23, SD = 0.76$), and Item 30 – *the structure of language* ($N = 96, M = 3.03, SD = 0.84$). These two items not only had the highest means, but also some of the smallest standard deviations, which emphasizes that the teachers concur that knowing about language and how language skills develop is an important knowledge base to have as a language teacher.

The participants endorsed four items that were coded as Language Pedagogy. Item 22 – *how assessments can influence teaching and learning in the classroom* ($N = 96, M = 3.09, SD = 0.81$); and Item 21 – *how assessments can influence teaching and learning materials* ($N = 96, M$

= 3.06, $SD = 0.78$). These two items addressed the language teachers' professional lives as language educators, and asked participants to recognize the relationship among classroom teaching and assessing. The participating Uzbekistan EFL teachers also identified Item 19 – *how to give useful feedback on the basis of assessment* ($N = 96, M = 3.05, SD = 0.80$) and Item 3 – *how to use assessments to evaluate achievement in language learning* ($N = 96, M = 3.04, SD = 0.78$) as important items for classroom language teachers. Similar to Items 23 and 30, Items 3, 19, 21, and 22 also had low standard deviations.

6.1.4.2 Items that were not Endorsed as Strongly

Based on Kremmel and Harding's original categorization, participants did not endorse as strongly items that addressed their personal beliefs/attitudes, using statistics, or communicating results to students and parents (see Table 18). Participants provided a mean score below 2.56 for Items 14, 41, 43, 45, and 46.

Table 18 Lowest Means

Item	Item Description	Mean	Median	Mode	SD
41	<i>one's own beliefs/attitudes towards language assessment</i>	2.55	3.00	3.00	0.91
46	<i>using statistics to analyze overall scores on a particular assessment</i>	2.53	2.00	2.00	0.95
45	<i>using statistics to analyze the difficulty of individual items (questions)</i>	2.52	2.00	2.00	1.00
14	<i>how to communicate assessment results and decisions to students or parents</i>	2.52	2.00	2.00	0.93
43	<i>how one's own beliefs/attitudes may conflict with those of other groups involved in assessment</i>	2.35	2.00	3.00	1.07

One item was coded as Scores and Decision Making. Item 14 – *how to communicate assessment results and decisions to students or parents* ($N = 96, M = 2.52, SD = 0.93$) had a lower mean (from all the means in the distribution) and a median and mode of 2.00. This result could suggest that the participants do not see their role as university English instructors to extend to communicating with parents.

Two items were coded under Personal Beliefs and Attitudes. Item 41 – *one’s own beliefs/attitudes towards language assessment* ($N = 96, M = 2.55, SD = 0.91$); and Item 43 – *how one’s own beliefs/attitudes may conflict with those of other groups involved in assessment* ($N = 96, M = 2.35, SD = 1.07$) were coded as Personal Beliefs/Attitudes and received the lowest mean scores from participants. Item 43’s median score of 2.00 also had one of the top two largest standard deviations.

Two items were coded under Technical Skills. Items 45 and 46 had similar means, medians (2.00), and modes (2.00). Item 46 – *using statistics to analyze overall scores on a particular assessment* ($N = 96, M = 2.53, SD = 0.95$); Item 45 – *using statistics to analyze the difficulty of individual items (questions)* ($N = 96, M = 2.52, SD = 1.00$) both discussed statistics. This result could suggest the teachers’ interest and expertise in mathematics is not strong.

6.1.5 A Visual Representation of Variability

Visual representations were created for Item 23 – *how language skills develop* ($N = 96, M = 3.23, SD = 0.76$) and Item 43 – *how one’s own beliefs/attitudes may conflict with those of other groups involved in assessment* ($N = 96, M = 2.35, SD = 1.07$) to illustrate the visual pattern on the extreme ends of the data (see Figure 13). Item 23 has the highest mean, smallest standard deviation, while Item 43 has the lowest mean and largest standard deviation.

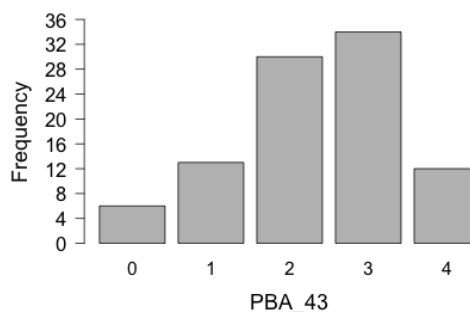
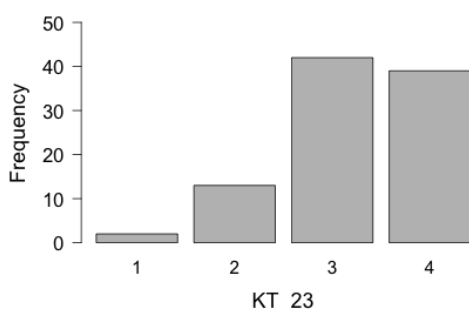


Figure 13 Histograms of Item 23 (KT_23) and Item 43 (PBA_43)

The histograms for Items 23 and 43 have negative skewness, indicating that the sampling distribution of the two items with the highest and lowest mean have similar patterns. The participants rated all items in the survey as somewhat important, and each item had a visual graphic representation similar to those of Items 23 and 43.

6.1.6 How did the participating Uzbekistan EFL teachers respond?

Participating Uzbekistan EFL teachers responded positively to each item presented to them in the Language Assessment Literacy Survey (Kremmel & Harding, *forthcoming*). The measures of dispersion, similar to the measures of central tendency, emphasize the little variability within and between the items in the data set, in which all item-level response patterns were negatively skewed. This result showed a trend of each item being at least somewhat important. Particularly, the Uzbekistan EFL teachers rated items of higher importance if they fell under the three factors (Kremmel & Harding, *forthcoming*) of Washback and Preparation; Language Structure, Use, and Development; and Assessment in Language Pedagogy. The three factors that received attention in the middle of importance were Assessment Principles and Interpretation, Scoring and Rating, and Developing and Administering Language Assessments. The bottom three factors that received relatively lower attention, but were still seen as important, were Statistical/Research Methods, Assessment Policy and Local Practice, and Personal Beliefs and Attitudes. More specifically, participants strongly endorsed items that addressed language, the relationship between teaching and assessing, and providing feedback. Participants did not endorse as strongly (but still favored positively) items that addressed their beliefs/attitudes, using statistics, or communicating results to students and parents. Generally, participants viewed each individual category and/or item as important or somewhat important for language teachers to know and do.

6.2 Representation of Taylor (2013)

As explained in Chapters 2 and 4, the purpose of the creation of the Language Assessment Literacy Survey was to empirically validate the dimensions in Taylor's (2013) language assessment literacy profile descriptors in order to discern if different stakeholders in language assessment have different needs. She emphasized that "It should be noted that the labelled dimensions on the eight axes (i.e. knowledge of theory, technical skills, etc.) *are hypothesized from the discussion of possible AL/LAL components across various papers in this special issue* [emphasis mine], while the values (i.e. 0–4) are hypothesized according to the different stages of literacy suggested by Pill and Harding" (p. 409). Her presentation of the eight dimensions is open for debate and discussion, particularly in how the constructs are operationalized to inform the creation of survey items. Thus, I will answer the following question: Is the theoretical basis for the survey (Taylor's 2013 framework of LAL) faithfully implemented in the survey? Prior to receiving Kremmel and Harding's coding scheme and factor analysis, I conducted an analysis which consisted of two stages: (1) creating my own operational definitions of the language assessment literacy dimensions [Taylor (2013)]; and, (2) coding survey items with my own operational definitions. Then, I compared my codes with Kremmel and Harding's initial coding scheme and their factor analysis.

6.2.1 Creating Operational Definitions of the LAL Dimensions

Figure 14 depicts the process I undertook to create definitions of Taylor's dimensions:

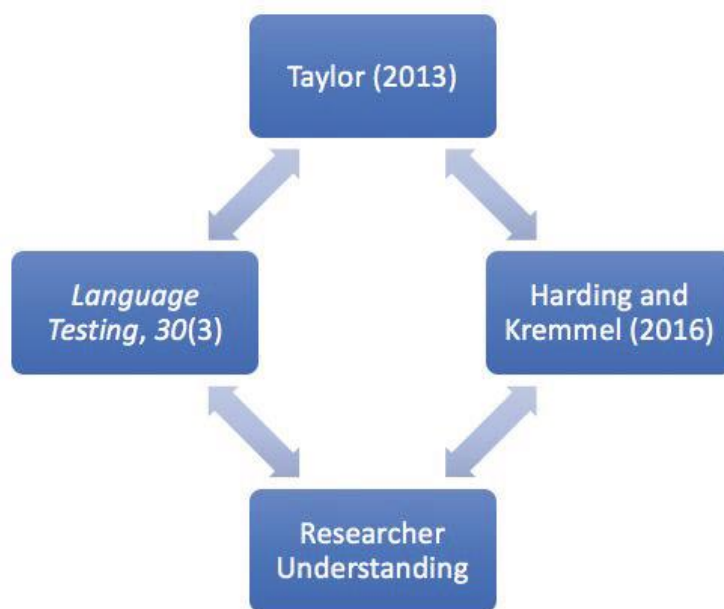


Figure 14 Researcher Understanding

First, I read through *Language Testing, 30(3)*, a special journal issue focusing on assessment literacy, and identified quotations that Taylor (2013) could have used to help her operationalize each of the eight constructs she hypothesized (see Appendix I). Second, I placed those quotations alongside the revised (expanded/divided) domains that Kremmel and Harding (2016) used for their *Language Assessment Literacy* (2016) poster presentation (see Appendix J for picture). Third, when I compared the quotations with the revised domain names side-by-side, I created my own operationalizations to code the Language Assessment Literacy survey items (see Table 19 for Researcher Operationalization).

Table 19 Researcher Operationalization

LAL Domains (Taylor, 2013)	Researcher Understanding Based on <i>Language Testing 30(3)</i> and Kremmel and Harding's (2016) Revised Domains.
Knowledge of Theory	<i>Knowledge of theory</i> is about theories of language, language learning, and the different stages of language development.
Technical Skills	<i>Technical skills</i> of assessment/testing include skills of test design/construction, administration, scoring, and test evaluation.
Principles and Concepts	<i>Principles and Concepts</i> are the fundamental concepts (both declarative and procedural knowledge) about language assessment and testing.
Language Pedagogy	<i>Language Pedagogy</i> is about teachers' actions, judgments, and teaching strategies.
Sociocultural Values	<i>Sociocultural Values</i> for language assessment literacy refers to the interrelationship between social, political and cultural values of a society and how language tests/assessments are constructed, administered, and interpreted from these cultural values.
Local Practices	<i>Local Practices</i> are context-specific assessment practices that are fixed to a local context/language program.
Personal Beliefs and Attitudes	<i>Personal Beliefs and Attitudes</i> refers to understandings of their own affective factor: tensions, frustrations, and/or happiness with assessments/testing.
Scores and Decision Making	<i>Scores and Decision Making</i> is about what assessment stakeholders <i>do</i> with the results of the assessments.

These eight definitions were written to connect to Taylor (2013), the language testing specialists' discussions of the AL/LAL construct throughout the special edition of *Language Testing 30(3)* (Jeong, Malone, O'Loughlin, Pill & Harding, and Scarino), and Kremmel and Harding's (2016) expanded LAL domain definitions.

6.2.2 Coding Survey Items

Figure 15 depicts the process I undertook to code the English version of the Language Assessment Literacy Survey, and then compare it with the survey creators' coding.

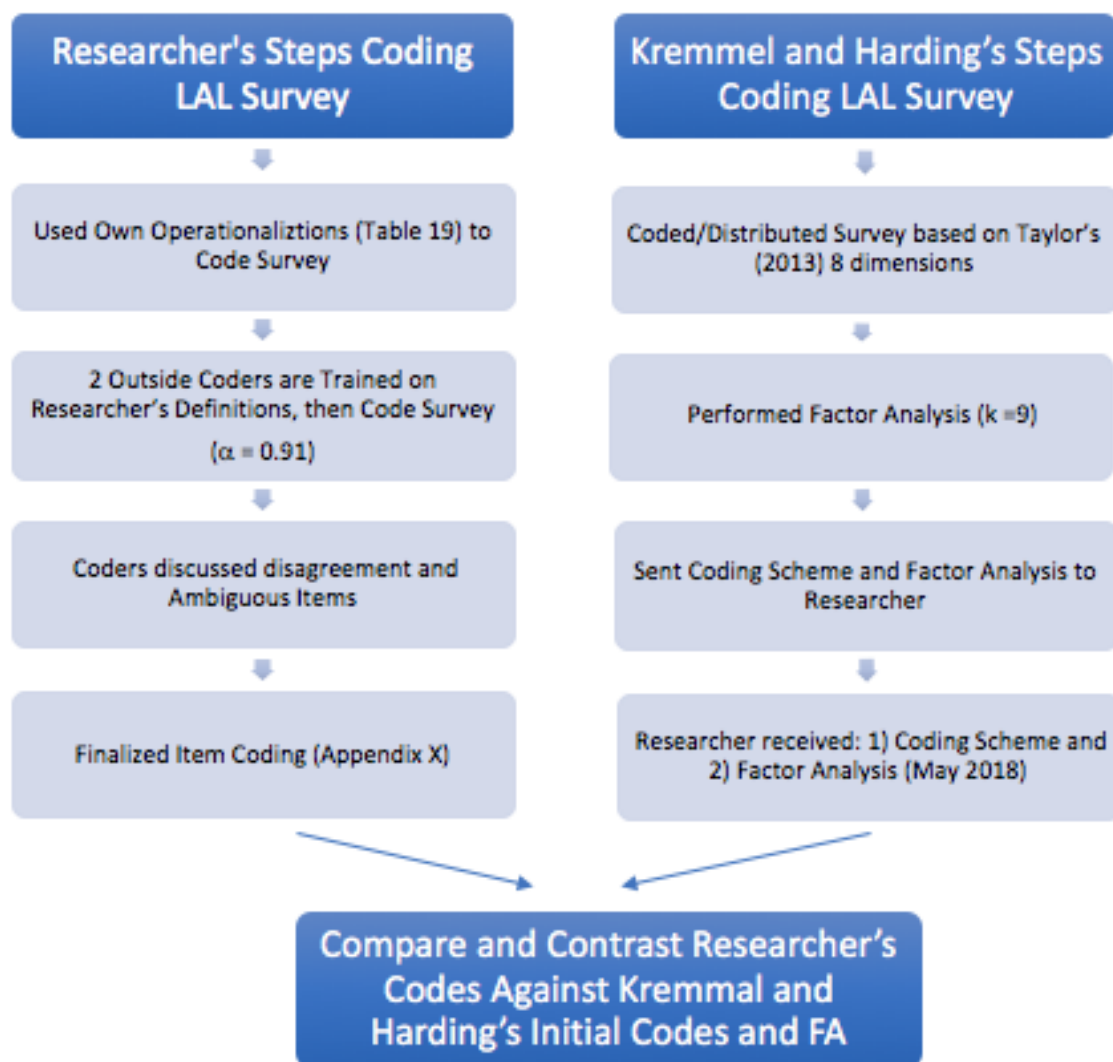


Figure 15 Coding LAL Survey

First, I read through each survey item, which were written to correspond to one of eight categories that Taylor (2013) hypothesized as part of the language assessment literacy construct: (1) knowledge of theory [KnofTh]; (2) technical skills [TeSk]; (3) principles and concepts [PrCon]; (4) language pedagogy [LanPed]; (5) sociocultural values [SocVal]; (6) local practices [LocPrac]; (7) personal beliefs/attitudes [PerBelAtt]; and (8) scores and decision-making [ScDeMa]. I coded each item, as exemplified in the brackets above, to one of these eight

categories. Second, two outside coders were trained on my definitions and then they coded the survey separately based on my coding scheme.

6.2.2.1 *Items of Disagreement Identified by Chiesa and Coders*

In total, three coders labeled items from the Language Assessment Literacy Survey, and inter-coder agreement ($\alpha = 0.91$) was achieved. After I calculated inter-coder agreement, I organized a group discussion and took notes on items that coders disagreed on (i.e., the codes that were not labeled the same way). Subsequently, coders emailed me if they felt they needed to clarify what was discussed during the focus group interview. There was disagreement on six survey items (i.e., 11, 13, 14, 18, 19, 60; see Table 20). I have placed a (*) next to the labels that Kremmel and Harding used for their original categorization. (I will compare my and Kremmel/Harding's codes in detail in Section 6.2.2.3.)

Table 20 Disagreement Among Coders

Item	Item Description	Coder 1	Coder II	Coder III
11	<i>how to determine if the content of a language assessment is culturally appropriate</i>	SocVal	LocPrac *	SocVal
13	<i>how to communicate assessment results and decisions to teachers</i>	ScDeMa*	SocVal	(-)
14	<i>how to communicate assessment results and decisions to students or parents</i>	ScDeMa*	SocVal	LanPed
18	<i>how to find information to help in interpreting results</i>	ScDeMa*	ScDeMa*	SocVal
19	<i>how to give useful feedback on the basis of assessment</i>	ScDeMa	LanPed*	ScDeMa
60	<i>aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])</i>	PrCon	TeSk*	PrCon

Note. SocVal = Sociocultural Values; LocPrac = Local Practices; ScDeMa = Scores and Decision Making, LanPed = Language Pedagogy; TeSk = Technical Skills; PrCon = Principles and Concepts; (-) = no coding; (*) = coded the same as Kremmel and Harding's original categorization.

Item 11 – *How to determine if the content of a language assessment is culturally appropriate* was identified by Coders One and Three as Sociocultural Values, while Coder Two labeled it as Local Practices. Coder One believed that “for a judgement to be made about an assessment content being culturally appropriate, one would need to be culturally competent and be able to identify and understand sociocultural values within a local context” (e-mail, October 2017). Thus, for this item, Sociocultural Values was identified by Coders One and Three as a superordinate topic to Local Practices. However, Coder Two disagreed and emphasized that a Local Practice is context specific and is a superordinate topic to Sociocultural Values.

Items 13 – *how to communicate assessment results and decisions to teachers*; 14 – *how to communicate assessment results and decisions to students or parents* were both coded as Scores and Decision Making from Coder One and Sociocultural Values from Coder Two. Coder Three did not provide a label for Item 13. The phrase – *how to communicate* – indicated to Coder Two to identify Items 13 and 14 as Sociocultural Values because communication/discourse is context and culture specific. For people to know *how to communicate*, they would need to be culturally competent. Coder Three did not label Item 13, but did identify Item 14 as Language Pedagogy because he believed that, “part of pedagogy is the teachers’ ability to communicate with students and (sometimes) parents” (e-mail, October 2017).

Item 18 – *how to find information to help in interpreting results* was labeled by Coders One and Two as Scores and Decision Making but Coder Three rated it as Sociocultural Value. Coder One explained that he thinks “The key word in this Item is on the word, interpreting” (personal communication, October 2017). How somebody interprets and communicates information appropriately in a context is dependent on the Sociocultural Value norms.

Item 19 – *how to give useful feedback on the basis of assessment* was coded as Scores and Decision Making by Coders One and Three and Language Pedagogy for Coder Two. This item was identified by the Coders as extremely difficult to distinguish because giving feedback is what teachers do inside and outside of their classrooms, and can be identified as language pedagogy. Knowing how to give *useful* feedback, however, could be dependent on one’s knowledge how to interpret assessments appropriately to make the right decisions, which meant to Coder One and Three to label Item 19 as Scores and Decision Making. Additionally, they believed that how one gives feedback could be context and culture specific.

Item 60 – *Aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])* – was coded as Principles and Concepts but could also be identified as a Technical Skill, if the emphasis of the item is on the participants knowing *how to align* tests with proficiency standards.

The six items discussed (i.e., 11, 13, 14, 18, 19, 60) caused disagreement among coders. I determined the final coding of each item based on which LAL descriptor received more attention (votes). For items 13 and 14 – which did not have a ‘majority’ coded category – I coded them based on my initial coding. Not only was there disagreement, but the coders also identified ambiguity of survey items. In other words, coders double-labeled items they felt fit into more than one dimension.

6.2.2.2 Ambiguous Items Identified by Chiesa and Coders

Of the 60 items that were coded the same, twelve items were acknowledged by the raters during the focus group discussion as ambiguous (i.e., another code was possible). See Table 21 for a breakdown of the items, their coding by raters, and potential other categories:

Table 21 Ambiguity of Items

Item	Item Description	Coded	Other Possible Coding
15	<i>how to determine whether a language assessment is useful for a particular purpose</i>	PrCon	TeSk
20	<i>how assessments can influence the design of a language course or curriculum</i>	LanPed	PrCon
21	<i>how assessments can influence teaching and learning materials</i>	LanPed	PrCon
22	<i>how assessments can influence teaching and learning in the classroom</i>	LanPed	LocPrac
31	<i>the advantages and disadvantages of standardized testing</i>	PerBelAtt	PrCon
32	<i>the philosophy behind the design of a relevant language assessment</i>	PrCon	KnofTh
44	<i>how one's own knowledge of language assessment might be further developed</i>	PerBelAtt	SocVal
49	<i>using rating scales to score speaking or writing performances</i>	TeSk	PrCon
56	<i>selecting appropriate items or tasks for a particular assessment purpose.</i>	PrCon	TeSk
61	<i>determining pass-fail marks or cut-scores</i>	TeSk	PrCon
62	<i>identifying assessment bias</i>	TeSk	PrCon
66	<i>selecting appropriate ready-made assessments.</i>	PrCon	TeSk

Twelve items (i.e., 15, 20, 21, 22, 31 32, 44, 49, 56, 61, 62, 66) were identified by the coders as ambiguous because each item could be identified as multiple categories. First, Item 15 – *how to determine whether a language assessment is useful for a particular purpose* – was coded as the category Principles and Concepts, but could also be identified as a Technical Skill or Local Practice. If the participant placed more emphasis on the “how to...”, then this procedural knowledge can be classified as a skill instead of only a concept. Item 20 – *how assessments can influence the design of a language course or curriculum* – and Item 21 – *how assessments can influence teaching and learning materials* – were initially coded as Language Pedagogy, but could also be identified as Principles and Concepts.

Item 44 – *how one's own knowledge of language assessment might be further developed* was coded as Personal Beliefs and Attitudes, but could also be identified as Sociocultural Values.

This item is especially vague because I am not sure I know what the item is trying to address. Were the survey developers trying to find out if the participants have a positive or negative attitude toward developing one's assessment practices? Or were they trying to ask if it is important for participants to know if they know different methods or ways to develop assessment skills? This item is particularly ambiguous.

Item 22 – *how assessments can influence teaching and learning in the classroom* was coded as Language Pedagogy because it is an item that directly discusses the classroom. However, it can also be coded as Local Practices, depending on how local practices are conceptualized within a classroom, school, or city (to give only some examples).

Item 31 – *the advantages and disadvantages of standardized testing* was coded as Personal Beliefs and Attitudes, but could also be identified as Principles and Concepts. Advantages and disadvantages are not objective ideas but are surrounded by attitudes and beliefs. Thus, to understand the advantages and disadvantages of standardized testing, participants would have to know what standardized testing entails, then their own histories and cultural backgrounds would play a role in choosing if one aspect is an advantage or disadvantage.

Item 32 – *the philosophy behind the design of a relevant language assessment* was identified as Principles and Concept but could also be Local Practices. The word *relevant* shows how the concepts of language assessment are tied to the local context.

Items 49 – *using rating scales to score speaking or writing performances*; 61 – *Determining pass-fail marks or cut-scores*; and 62 – *Identifying assessment bias* were initially coded as Technical Skills, but could also be coded as Principles and Concepts. To determine marks or identify bias, participants would need to know what these academic concepts are before they can do the *determining* or *identifying*.

Items 56 – *selecting appropriate items or tasks for a particular assessment purpose*; and 66 – *selecting appropriate ready-made assessments* were identified as Principles and Concepts but could also be Technical Skills. Whether the participants who take the survey understand the item as a *skill* or a *principle* can determine how they see the importance of it.

The twelve items (15, 20, 21, 22, 31 32, 44, 49, 56, 61, 62, 66) were identified by the coders as ambiguous because each item could be identified as belonging to multiple categories. I determined the final coding of each item based on the original codes that everyone agreed on. (See Appendix K for preliminary codings by Chiesa and two additional coders.) Next I compared my coded categories with the survey author’s coded items.

6.2.2.3 Comparison: Chiesa and Kremmel/Harding’s Initial Coding

After I received Kremmel and Harding’s initial categorization of each item, I calculated inter-coder agreement ($\alpha=.79$) between my final coding and theirs. Fifty-one of 66 items were coded similarly, while 15 of 66 items were coded differently. I will present the differences below. (See Tables 22, 23, and 24 for differences.)

Table 22 Chiesa and Kremmel/Harding Coding Comparison I

Item	Item Description	Chiesa	Kremmel/Harding
1	<i>how to use assessments to inform learning or teaching goals</i>	ScDeMa	LangP
2	<i>how to use assessments to evaluate progress in language learning</i>	ScDeMa	LangP
3	<i>how to use assessments to evaluate achievement in language learning</i>	ScDeMa	LangP
4	<i>how to use assessments to diagnose learners’ strengths and weaknesses</i>	ScDeMa	LangP
5	<i>how to use assessments to motivate student learning</i>	ScDeMa	LangP
6	<i>how to use self- and peer-assessment</i>	ScDeMa	LangP
16	<i>how to recognize when an assessment is being used inappropriately</i>	ScDeMa	ISV
19*	<i>how to give useful feedback on the basis of assessment</i>	ScDeMa	LangP

Note. Taylor’s (2013) category Sociocultural Values was renamed “Impact and Sociocultural Values” (ISV). The (*) symbol indicates an item disagreed upon from Section 5.2.2.

Eight items were coded by me as Scores and Decision Making (1, 2, 3, 4, 5, 6, 16, 19) while seven of those items were identified by Kremmel and Harding as Language Pedagogy (1, 2, 3, 4, 5, 6, 19). Item 16 – *how to recognize when an assessment is being used inappropriately* was labeled as Scores and Decision Making by me but as Impact and Sociocultural Values by Kremmel and Harding.

I coded five Items as Principles and Concepts (i.e., Items 27, 32, 56, 60, 66) (see Table 23 below.) These five items were labeled differently by Kremmel and Harding.

Table 23 Chiesa and Kremmel/Harding Coding Comparison II

Item	Item Description	Chiesa	Kremmel/Harding
27	<i>how pass-fail marks or cut-off scores are set</i>	PrCon	SDM
32	<i>the philosophy behind the design of a relevant language assessment</i>	PrCon	ISV
56	<i>selecting appropriate items or tasks for a particular assessment purpose</i>	PrCon	TS-A
60*	<i>aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference)</i>	PrCon	TS-C
66	<i>selecting appropriate ready-made assessments.</i>	PrCon	TS-A

Note. Taylor's (2013) category, *Technical Skills* was expanded into three domains: (TS-A) *Language assessment construction*; (TS-B) *Language assessment administration/scoring*; (TS-C) *Language assessment evaluation*. The (*) symbol indicates an item that was disagreed upon from Section 5.2.2.

Item 27 – *how pass-fail marks or cut-off scores are set* – was labeled as Scores and Decision Making; Item 32 – *The philosophy behind the design of a relevant language assessment* – was coded as Impact and Sociocultural Values. There were three Items (56, 60, 66) that were labeled by me as Principles and Concepts but were identified as Technical Skills. More specifically, Item 56 – *selecting appropriate items or tasks for a particular assessment purpose*; and, Item 66 – *selecting appropriate ready-made assessments* – were coded as Technical Skills (A) Language Assessment Construction. Item 60 – *Aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference)* – was identified as Technical Skills (C) Language Assessment Evaluation.

Two items (i.e., 11, 31) presented difficulties in the original coding, and were also identified as issues here. The arguments presented above are repeated here in the differences between the labeling of items by Chiesa and Kremmel/Harding (see Table 24).

Table 24 Chiesa and Kremmel/Harding Coding Comparison III

Item	Item Description	Chiesa	Kremmel/Harding
11*	<i>how to determine if the content of a language assessment is culturally appropriate</i>	SocVal	LocP
31*	<i>the advantages and disadvantages of standardized testing</i>	PerBelAtt	ISV

Note. The (*) symbol indicates an item that was disagreed or found to be ambiguous from Section 5.2.2

Item 11 – *how to determine if the content of a language assessment is culturally appropriate* – was labeled as Sociocultural Values by me and as Local Practices by Kremmel and Harding.

This distinction is similar with the initial disagreement amongst the coders (see Section 5.2.2).

Item 31 – *the advantages and disadvantages of standardized testing* – was identified as Personal Beliefs and Attitudes by the researcher, but then, coded as Impact and Sociocultural Values by Kremmel and Harding. This distinction reflects the ambiguity the coders identified in Section 6.2.2.2.

6.2.3 Summary of Initial Findings

Thirty-one of 66 items of the Language Assessment Literacy Survey have either been (1) disagreed on among coders ($k = 6$), (2) identified as ambiguous ($k = 12$), or (3) contained different codings from survey authors ($k = 13$). To identify how these items play a role in Kremmel and Harding's (*forthcoming*) final outcome of survey items, I examined the relationship between the problematic items and their factor analysis. This analysis would inform me if these items are considered 'unacceptable' by Kremmel and Harding's factor analysis.

6.2.4 Compare: Chiesa's Coding with Kremmel/Harding's Factor Analysis

After I received Kremmel and Harding's factor analysis (see Table 25), I analyzed how the questionable items from my initial coding and inter coder agreement with the survey authors, played a role (or did not) in the results. [Their factor analysis ended up with 50 items, and the factor names lined up well with their initial coding, but not flawlessly (e.g., Factor 6 – Assessment Principles and Interpretation – contained coding from Principles and Concepts, Language Pedagogy, and Scores and Decision Making).]

Table 25 LAL survey: 9 Factor solution (50 items)(Kremmel & Harding, forthcoming)

Factor	Factor Name	k
1	Developing and Administering Language Assessments	14
2	Assessment in Language Pedagogy	6
3	Assessment Policy and Local Practices	6
4	Personal Beliefs and Attitudes	4
5	Statistical / Research Methods	4
6	Assessment Principles and Interpretation	4
7	Language Structure, Use, and Development	5
8	Washback and preparation	4
9	Scoring and Rating	3

Factor 1– Developing and administering language assessments – contains 3 Items that were identified as disagreements (Items 27, 56, 60) between Chiesa and Kremmel/Harding. Item 27 – *How pass-fail marks or cut-off scores are set*; Item 56 – *selecting appropriate items or tasks for a particular assessment purpose*; and, Item 60 – *Aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference)* were initially identified by Chiesa as Principles and Concepts. However, these items combined with others that addressed developing and administering language assessments. This relationship showed that the focus of the items is on the initial verb phrase, thus being a skill, and not a concept. This result clarifies how Kremmel and Harding were conceptualizing Technical Skills in writing of survey items.

Factor 2 – Assessment in Language Pedagogy – contains six items that were identified as disagreements between Chiesa and Kremmel/Harding (i.e., Items 1, 4, 5, 6, 19). [Item 6 – *how to use self- and peer-assessment* – was divided into two Items: (1) How to use self-assessment, and (2) How to use peer-assessment.] The disagreement lies in the *definitions* between the original categories of Scores and Decision Making and Language Pedagogy. Thus, from Kremmel and Harding’s labeling of Factor 2, and without a label of Scores and Decision Making in their Factor Analysis as originally in Taylor’s (2013) framework, we can surmise that *how* language teachers use scores and decision making *is a part of* Language Pedagogy. Thus, Kremmel and Harding’s Factor labeling cleared up the ambiguity in the definitions of these two categories I originally had with my coding team. Now, we have a better understanding of how Kremmel and Harding operationalized their constructs.

Factor 3 – Assessment Policy and Local practice; Factor 4 – Personal Beliefs and Attitudes; Factor 5 – Statistical / Research Methods; Factor 6 – Assessment Principles and Interpretation; Factor 7 – Language Structure, Use, and Development; Factor 8 – Washback and Preparation; and, Factor 9 – Scoring and Rating do not contain any items that were identified as problematic (i.e., ambiguous or causing disagreement) from my initial coding and comparison with survey authors. Appendix L presents the final coding scheme.

6.2.5 Representation of Taylor’s (2013) Framework?

With reference to the design of the Language Assessment Literacy survey, which includes operationalization of key constructs and the writing of items, I can conclude that Taylor’s (2013) framework lends itself to multiple interpretations. With my initial operationalizations of the language assessment literacy dimensions that Taylor (2013) identified, I have showed how different researchers perceived items as belonging to different categories (e.g., Scores and

Decision Making and Language Pedagogy). This aspect of the theoretical framework led to ambiguity of definitions and written items. Sometimes there were disagreements about which items corresponded to which category, and some items were identified as ambiguous (i.e., coders could not confidently place an item into a single category). However, Kremmel and Harding produced enough items, and had enough people ($N > 1000$) complete the English version of their survey to produce a meaningful set of dimensions and items that resolve problematic issues in the English version.

6.3 Factor Analysis

After selecting a coding scheme for the LAL subconstructs survey items were theorized to be assessing, I ran an Exploratory Factor Analysis (EFA) to answer the research question, “What is the underlying factor structure of the Language Assessment Literacy Survey?” Factor analysis is commonly used in the fields of psychology and education for interpreting self-reported questionnaires. Factor analysis does not consist of one statistical method but a variety of complex structure analyzing procedures, and is typically used to reduce a large number of variables into a smaller set of factors that are conceptually similar. Factor analysis establishes underlying dimensions between the measured variables (e.g., the scores given by the participants on survey items) and latent constructs (also known as its latent variables), thereby allowing the formation and refinement of theory and scale development.

Conducting a factor analysis provides construct validity evidence of self-reporting scales. For instance, if Taylor’s (2013) Language Assessment Literacy profile descriptor categories were clearer in their operationalizations, and thus, Kremmel and Harding’s Language Assessment Literacy survey items had been written to be less abstract given the context of interest, I would conduct a confirmatory factor analysis (CFA). A CFA would determine if there were indeed

eight different underlying constructs in the survey. However, the Language Assessment Literacy survey is in its nascent stage and the definitions of the constructs and the items have not been validated in certain contexts. Therefore, I ran an EFA to explore how the participating Uzbekistan EFL teachers responded to the survey.

According to Loewen and Gonulal (2015) there are two different types of Exploratory Factor Analysis (see Figure 16, second and third level). The first is called a Principal Component Analysis (PCA) and the other is an Exploratory (EFA).

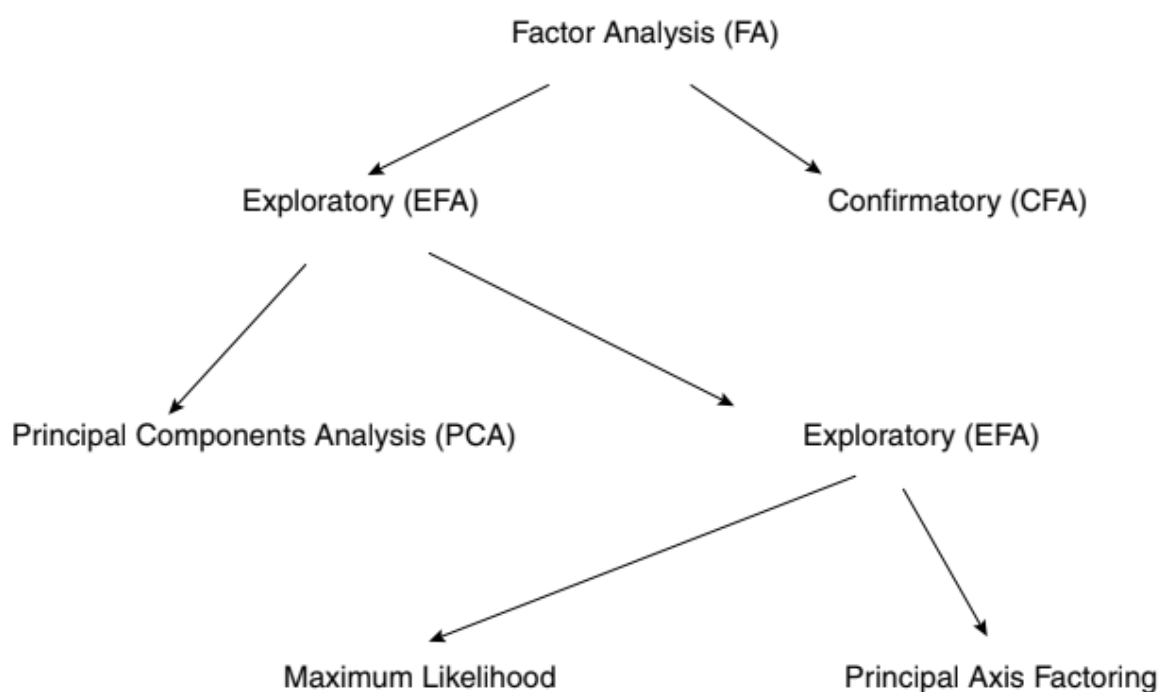


Figure 16 Loewen and Gonulal (2015), p. 184. Types of Factor Analyses

The difference between a PCA and EGA lies in that a PCA technique is used when you want to reduce the number of variables (for example, to make a long survey shorter) and an EFA is used when you want to understand the underlying structure of the survey (i.e., which survey items pattern together). Conway and Huffcut (2003) explain,

If a researcher's purpose is to understand the structure of a set of variables (which will usually be the case), then use of a common factor model [EFA] such as principal axis or maximum likelihood factoring represents a high-quality decision. If a researcher's purpose is pure reduction of variables... then the use of PCA represents a high-quality decision. (pp. 150-151)

Thus, I ran an EFA as opposed to a PCA because I wanted to understand the underlying structure of the set of variables in the Language Assessment Literacy Survey that I distributed to the Uzbekistan EFL teachers. Rietveld and Van Hout (1993) identified seven steps needed to conduct a factor analysis: (1) determining factorability of data; (2) choosing a factor extraction method; (3) deciding how many factors to retain (achieving simple structure); (4) selecting a factor rotation method; (5) determining factor loadings; (6) interpreting results; and (7) reporting results.

6.3.1 Determining Factorability of Data

The factorability of the data refers to whether the data set is strong enough, or contains enough information, for a factor analysis to be run. Loewen and Gonulal (2015) emphasize that the data set needs to be interval-like, including Likert scales (p. 187). A general rule of thumb is that if a researcher is working with people, then there needs to be a minimum of 100 people (Hair, Anderson, Tatham & Black, 1995). Because the number of participating language teachers was less than 100 ($N = 96$), I conducted a post-hoc analysis called the Kaiser-Meyer-Olkin (KMO) test to determine whether the data were suitable to run a factor analysis. KMO specifically measures sampling adequacy for each variable in the model and for the complete model. The statistical test helps determine suitability for factor analysis because it considers the proportion of variance of each variable with the total model. The lower the proportion, the more suited the

data is to conduct a Factor Analysis. The KMO value for the sample was .76, above the recommended value of 0.6 (Kaiser, 1974), suggesting that the data were appropriate for factor analysis.

6.3.2 *Choosing a Factor Extraction Method*

Based on the suitability of the data, and my purpose to find the latent variables in the survey based on Taylor's (2013) categories, I computed an Exploratory Factor Analysis (EFA) with the extraction method of *principal axis factoring*. The principal axis factoring is meant to recover weak factors and is preferred when attempting to identify latent constructs. To explore the factor structure of the Language Assessment Literacy Survey, I used JASP v.0.8.3.1 (Wagenmakers, 2018), which is a free and open-source graphical program for statistical analyses. It is written in the computer languages of C++, R, and Javascript.

6.3.3 *Deciding How Many Factors to Retain and Selecting a Factor Rotation*

The third and fourth steps – Deciding How Many Factors to Retain and Selecting a Factor Rotation – worked simultaneously in JASP. When I ran the EFA in JASP, I first conducted a principal axis factoring method with an oblique rotation (the default setting). From this initial analysis, three Factors were extracted with 3 items loading onto more than one factor. I then checked the scree plot (see Figure 17) to graphically identify the number of factors that should be extracted. The cut off point for selecting factors is the point of inflexion, which is the sharp descent, or elbow, in the slope of the plot. Scree plots can be difficult to interpret because determining the point of inflexion is subjective. For instance, Figure 17 shows that the point of inflexion could be on the third or fourth factor. Loewen and Gonulal (2015) suggest that because of its subjective nature, “it is useful to interpret the scree plot in light of other factor retention criteria” (p. 196).

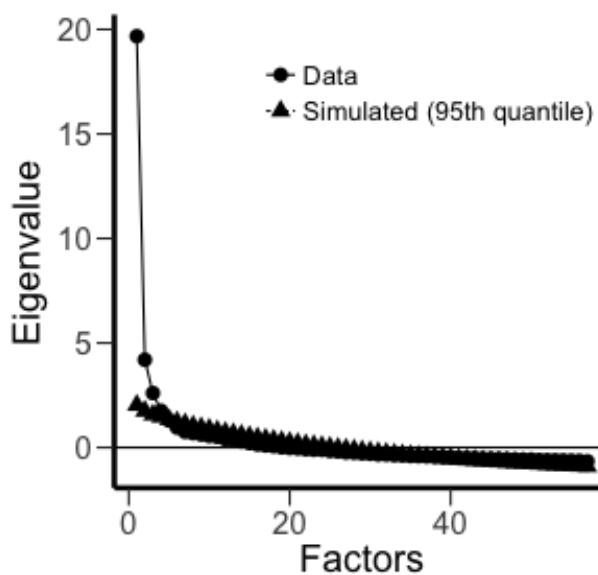


Figure 17 Scree Plot

Therefore, I decided to proceed with the initial analysis of the three factors extracted and keep the oblique rotation method instead of the orthogonal method rotation. Orthogonal rotation methods carry an assumption that the factors are uncorrelated, while in contrast, the oblique rotation carries an assumption that the factors are correlated. I checked for correlations among the factors and found that each one exceeded .32 (see Table 26 for correlation matrix).

Tabachnick and Fidell (2013) noted, “If correlations exceed .32, then there is 10% (or more) overlap in variance among factors, enough variance to warrant oblique rotation unless there are compelling reasons for orthogonal rotation” (p. 646).

Table 26 Correlation Matrix

	<u>Factor 1</u>	<u>Factor 2</u>	<u>Factor 3</u>
Factor 1	1.000		
Factor 2	0.573	1.000	
Factor 3	0.538	0.627	1.000

Once I determined that there were correlations among the factors, I began removing the double-loaded items to help me decide how many factors to retain for a simple structure. Double-

loaded items meant that one item loaded onto two factors. I then removed an item (or multiple items) from analysis, and then JASP simultaneously recalculated the EFA. With each successive output of EFA, there were three to four items that loaded onto more than one Factor. More specifically, three items were removed after the first round, four items were removed after the second round, and two items were removed after the third round. After three rounds, I had removed a total of nine items (i.e., PerBeAtt_31; PrCon15, PrCon_36, PrCon_39, PrCon_60, KnofTh_37, KnofTh_38, TeSk_45, and, Tesk_46), and a simple structure with three Factors was achieved – in other words, only one survey item corresponded to each single factor. The removed items are in Table 27.

Table 27 Items Removed from Factor Analysis to Achieve a More Parsimonious Structure

Item	Item Description
37	<i>different language proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])</i>
39	<i>different types and purposes for language assessment (e.g., proficiency, achievement, diagnostic)</i>
38	<i>different stages of language proficiency</i>
60	<i>aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference)</i>
15	<i>how to determine whether a language assessment is useful for a particular purpose.</i>
36	<i>the specialist terminology related to language assessment.</i>
31	<i>the advantages and disadvantages of standardized testing</i>
46	<i>using statistics to analyze overall scores on a particular assessment</i>
45	<i>using statistics to analyze the difficulty of individual items (questions)</i>

To identify if the simple structure achieved is an adequate solution, I checked statistical model fit. The Tucker-Lewis Index, evaluation of goodness of fit, was 0.91. Acceptable fit is above 0.90 (In'nami & Koizumi, 2015).

6.3.4 Determining Factor Loadings

Three factors with their loadings are presented in Table 28 below. Factor 1 contained nineteen loadings; Factor 2 had sixteen loadings; Factor 3 had twenty-two loadings.

Table 28 Factor Loadings

Item Code	Factor 1	Factor 2	Factor 3	Survey Items
TeSk_54	0.877	-0.262	0.069	<i>developing specifications (overall plans) for language assessment</i>
TeSk_49	0.818	-0.214	0.141	<i>using rating scales to score speaking or writing performances</i>
TeSk_47	0.808	0.094	-0.240	<i>using statistics to analyze the quality of individual items (questions)/tasks</i>
TeSk_59	0.757	0.004	0.020	<i>writing good quality items (questions) or tasks for language assessments</i>
TeSk_48	0.755	-0.042	-0.076	<i>using techniques other than statistics (e.g., questionnaires, interviews, analysis of language) to get information about the quality of language assessment</i>
PrCon_64	0.745	-0.003	0.069	<i>making decisions about what aspects of language to assess</i>
PrCon_65	0.741	0.050	-0.095	<i>piloting/trying-out assessments before their administration</i>
TeSk_55	0.735	0.025	-0.044	<i>selecting appropriate rating scales (rubrics)</i>
TeSk_58	0.720	-0.042	0.042	<i>training others to write good quality items (questions) or tasks for language assessment</i>
TeSk_50	0.686	0.190	-0.088	<i>using specifications to develop items (questions) and tasks</i>
TeSk_51	0.670	-0.083	0.183	<i>scoring closed-response questions (e.g., Multiple Choice Questions)</i>
TeSk_52	0.667	-0.072	0.168	<i>scoring open-ended questions (e.g., short answer questions)</i>
TeSk_57	0.660	-0.113	0.202	<i>training others to use rating scales (rubrics) appropriately</i>
TeSk_63	0.646	0.084	0.039	<i>designing scoring keys and rating scales (rubrics) for assessment tasks</i>
TeSk_61	0.623	0.174	-0.038	<i>determining pass-fail marks or cut-scores</i>
PrCon_56	0.622	0.142	-0.021	<i>selecting appropriate items or tasks for a particular assessment purpose</i>
TeSk_62	0.612	0.263	-0.140	<i>identifying assessment bias.</i>
PrCon_66	0.575	0.223	0.019	<i>selecting appropriate ready-made assessments.</i>
TeSk_53	0.563	0.096	-0.053	<i>developing portfolio-based assessments</i>
PerBelAtt_42	0.090	0.829	-0.233	<i>how one's own beliefs/attitudes might influence one's assessment practices</i>
PerBelAtt_41	-0.084	0.817	0.023	<i>one's own beliefs/attitudes towards language assessment</i>
PrCon_32	-0.071	0.761	0.029	<i>the philosophy behind the design of a relevant language assessment</i>

Item Code	Factor 1	Factor 2	Factor 3	Survey Items
PrCon_40	-0.129	0.731	0.027	<i>different forms of alternative assessments (e.g., portfolio assessment)</i>
SocVal_26	0.050	0.696	-0.041	<i>how social values can influence language assessment design and use</i>
LocPrac_35	0.185	0.667	-0.115	<i>the assessment traditions in a local context</i>
SocVal_25	0.065	0.661	-0.111	<i>how language is used in society</i>
PrCon_29	-0.036	0.660	0.140	<i>the concept of validity</i>
SocVal_33	-0.035	0.634	0.120	<i>the impact language assessments can have on society</i>
PerBelAtt_44	0.196	0.623	-0.108	<i>how one's own knowledge of language assessment might be further developed</i>
LocPrac_34	0.232	0.622	-0.067	<i>the relevant legal regulations for assessment in the local area</i>
PrCon_27	-0.052	0.543	0.234	<i>how pass-fail marks or cut-off scores are set</i>
PerBelAtt_43	0.262	0.531	-0.029	<i>how one's own beliefs/attitudes may conflict with those of other groups involved in assessment</i>
PrCon_28	-0.084	0.530	0.267	<i>the concept of reliability</i>
ScDeMA_14	-0.033	0.461	0.192	<i>how to communicate assessment results and decisions to students or parents</i>
ScDeMA_18	-0.028	0.403	-0.147	<i>how to find information to help in interpreting results</i>
ScDeMA_04	0.094	-0.136	0.754	<i>how to use assessments to diagnose learners' strengths and weaknesses</i>
LanPed_21	-0.054	-0.038	0.731	<i>how assessments can influence teaching and learning materials</i>
ScDeMA_19	-0.065	-0.050	0.723	<i>how to give useful feedback on the basis of assessment</i>
ScDeMA_01	-0.059	0.092	0.706	<i>how to use assessments to inform learning or teaching goals</i>
ScDeMA_03	-0.029	-0.071	0.702	<i>how to use assessments to evaluate achievement in language learning</i>
KnOfTh_23	-0.173	0.083	0.684	<i>how language skills develop</i>
ScDeMA_07	0.102	-0.200	0.633	<i>how to interpret assessment results appropriately</i>
SocVal_11	0.318	-0.286	0.624	<i>how to determine if the content of a language assessment is culturally appropriate</i>
ScDeMA_16	-0.148	0.059	0.599	<i>how to recognize when an assessment is being used inappropriately</i>
KnOfTh_24	-0.143	0.190	0.585	<i>how foreign/second languages are learned</i>
ScDeMA_02	0.081	0.059	0.562	<i>how to use assessments to evaluate progress in language learning</i>

Item Code	Factor 1	Factor 2	Factor 3	Survey Items
ScDeMA_05	0.038	0.002	0.560	<i>how to use assessments to motivate student learning</i>
KnOfTh_30	-0.011	0.186	0.556	<i>the structure of language</i>
ScDeMa_06	0.021	0.002	0.516	<i>how to use self- and peer-assessment</i>
LanPed_17	-0.063	0.240	0.511	<i>how to prepare learners to take language assessments</i>
ScDeMA_13	-0.099	0.340	0.493	<i>how to communicate assessment results and decisions to teachers</i>
LanPed_22	0.017	0.240	0.451	<i>how assessments can influence teaching and learning in the classroom</i>
ScDeMA_08	0.237	-0.020	0.447	<i>how to interpret measurement error</i>
LanPed_20	0.026	0.148	0.430	<i>how assessments can influence the design of a language course or curriculum</i>
LocPrac_12	0.315	0.011	0.423	<i>how to determine if the results of the assessment are relevant to the local context</i>
LocPrac_10	-0.014	0.250	0.417	<i>how to determine if a local assessment aligns with a local education system</i>
ScDeMA_09	0.158	0.046	0.412	<i>how to interpret what a particular score says about an individual's language ability</i>

Note. KnOfTh = Knowledge of Theory; LanPed = Language Pedagogy; LocPrac = Local Practices; PerBelAtt = Personal Beliefs/Attitudes; PrCon= Principles and Concepts; ScDeMA = Scores and Decision Making; SocVal= Sociocultural Values; and, TeSk = Technical Skills.

6.3.5 Interpretation

The EFA uncovered three categories, which I have named: Technical Skills of Language Assessment, Social Aspects of Language Assessments, and Assessment-for-Learning. To name each category, Field (2009) recommended that one choose the names based on the five items loading most highly on each factor.

The first factor (Technical Skills of Assessment – TSA) represents a group of items that addresses putting technical skills of assessment into practice. The top five loadings address how to: (1) use statistics to analyze the quality of items; (2) use rating scales to score speaking and writing performance; (3) develop specifications for language assessments; (4) use techniques other than statistics to get information about the quality of language; and (5) write high-quality

items or tasks for language assessments. The remaining fourteen items in this factor also address specific skills of assessment that a person would need to be able to possess.

The second factor (Social Aspects of Language Assessments – SALA) represents a group of items that address the social/personal sphere about language assessment. The top five loadings address knowing: (1) one's own beliefs/attitudes about language assessment; (2) how one's own beliefs/attitudes might influence one's assessment practices; (3) the philosophy behind the design of a relevant language assessment; (4) different forms of alternative assessments; and (5) how social values can influence language assessment design and use. The remaining eleven items in this factor also address the social/personal aspect of assessment.

The third factor is Assessment-*for*-Learning (AfL), which is often referred to as those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used to modify the teaching and learning activities in which they are engaged (Black & William, 1998). The top five loadings in this factor are the following: (1) how to use assessments to diagnose learners' strengths and weaknesses; (2) how to give useful feedback on the basis of assessments; (3) how to use assessments to inform learning or teaching goals; (4) how to use assessments to evaluate achievement in language learning; and (5) how to interpret assessment results appropriately. The remaining seventeen items in this factor are about the Assessment-*for*-Learning construct and include aspects such as the structure of language, development of language skills, and how to communicate results to parents and students.

6.3.6 Reporting the Results

Table 29 presents the descriptive statistics for each extracted Factor and is ordered from the largest mean to the smallest.

Table 29 Descriptive Statistics – Factors

<u>Factor</u>	<u>Factor Name</u>	<u>Mean</u>	<u>Standard Deviation</u>	<u>Min</u>	<u>Max</u>	<u>Range</u>	<u>α</u>
Factor 3	AfL	2.87	0.17	2.57	3.23	0.66	.93
Factor 1	TSA	2.78	0.08	2.61	2.97	0.36	.93
Factor 2	SALA	2.66	0.14	2.35	2.89	0.54	.94

On average, Factor 3 (*AfL*) had the highest reported mean, the highest mean item standard deviation, the largest range (0.66), and a strong reliability ($\alpha = .93$). Thus, participants think somewhat similarly and perceive *AfL* as the most important knowledge base for language teachers to have. Factor 1 (Technical Skills of Assessment) had the second highest mean, lowest item SD, and strong reliability ($\alpha = .93$). Participants identified the technical skills of assessment the most similarly, and believe that technical skills are somewhat important for language teachers to be able to know about and do. Factor 2 (Social Aspects of Language Assessment) had the lowest reported mean, second highest item mean standard deviation, and a strong reliability ($\alpha = .94$). This result suggests there was a relatively wide range of opinions about the importance of the social aspect of assessment knowledge and skills among participants, where many language teachers do not see its importance to the language teaching profession. *Assessment-for-Learning* was thus perceived by the participants as slightly more important than Technical Skills of Assessment and Social Aspects of Language Assessment.

6.3.7 Chiesa EFA and Kremmel / Harding EFA Comparison

There is a relationship in the factor structure of Chiesa's EFA and Kremmel and Harding's EFA. (See Figures 18 and 19 for visual representations of the relationship between them.)

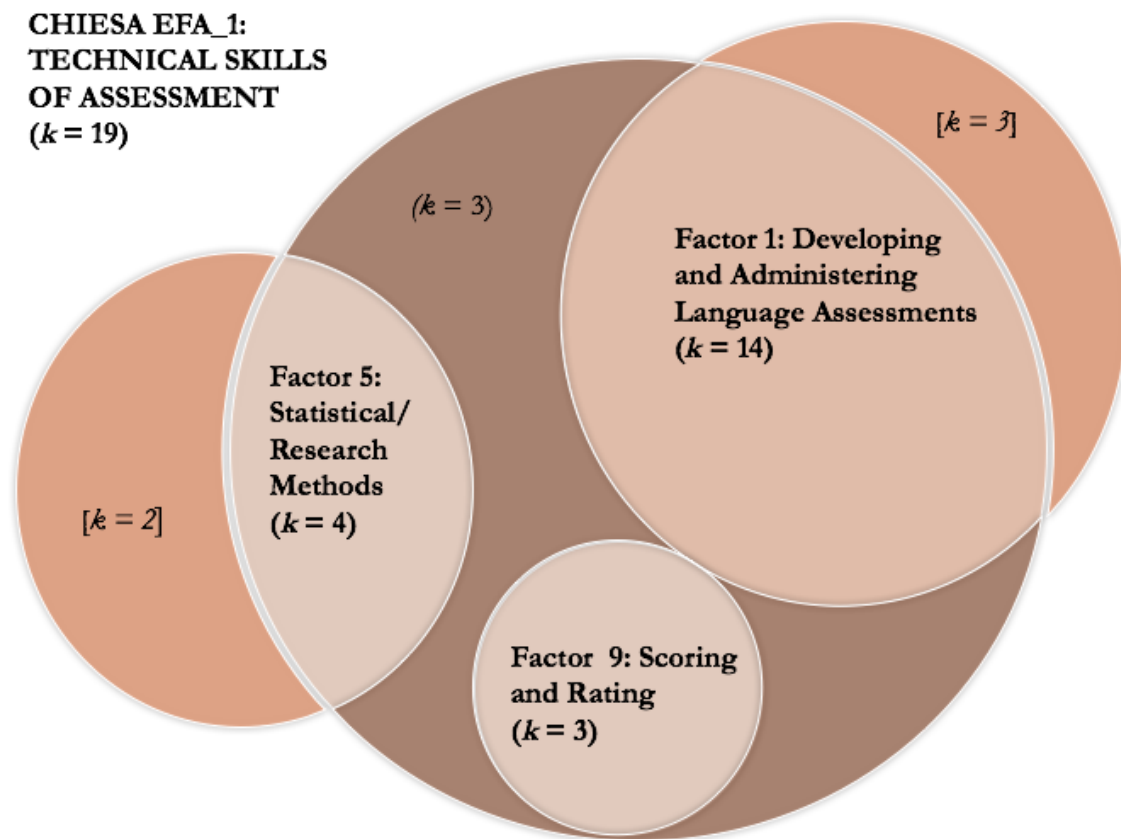


Figure 18 Chiesa, Factor 1: Technical Skills of Assessment w/ Kremmel and Harding's EFA

Note. The [] indicates the number of items in K&H's factor that were not included in Chiesa EFA (e.g., K&H's first Factor contained 14 items. Of these, 11 items were also in Chiesa's first Factor, and 3 were only in K&H.)

Sixteen of 19 items overlap with Chiesa's Factor 1 – Technical Skills of Assessment, and Kremmel and Harding's Factor 1 – Developing and Administering Language Assessments; Factor 5 – Statistical / Research Methods; and, Factor 9 – Scoring and Rating. However, three items were identified in Chiesa's Factor 1 but not in Kremmel and Harding's EFA: Item 50 – *using specifications to develop items (questions) and tasks*; Item 53 – *developing portfolio-based assessments*; and, Item 66 – *selecting appropriate ready-made assessments*. Also, three items from Kremmel and Harding's Factor 1 were not in Chiesa's first factor and were all items I removed: [Item 60 – *aligning tests to proficiency frameworks (e.g., the Common European*

Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL]); Item X – Accommodating candidates with disabilities or other learning impairments; Item X – How to train others about language assessment.]

Additionally, two items were a part of Harding and Kremmel's Factor 5 and not Chiesa's Factor 1: Item 45 – *using statistics to analyze the difficulty of individual items*; and, Item 46 – *using statistics to analyze overall scores on a particular assessment*. All items in Kremmel and Harding's Factor 9 were a part of Chiesa's Factor 1. These results suggest there is a relationship between Chiesa's Factor 1 and Kremmel and Harding's Factors 1, 5, and 9. Kremmel and Harding's results shows the nuanced nature of the different types of technical skills one would need to be language assessment literate. Additionally, the close relationship between Chiesa and Kremmel/Harding shows that the participants in both studies were responding similarly to survey items that proposed technical skills of language assessment. Also, Kremmel and Harding's larger sample ($N > 1000$) might have allowed for more factors than my study.

Ten of 16 survey items overlap with Chiesa's Factor 2 – Social Aspects of Assessment, and Kremmel and Harding's Factor 3 – Assessment Policy and Local Practices; Factor 4 – Personal Beliefs and Attitudes; Factor 6 – Assessment Principles and Interpretation; and, Factor 7 – Language Structure, Use, and Development (see Figure 19 below.)

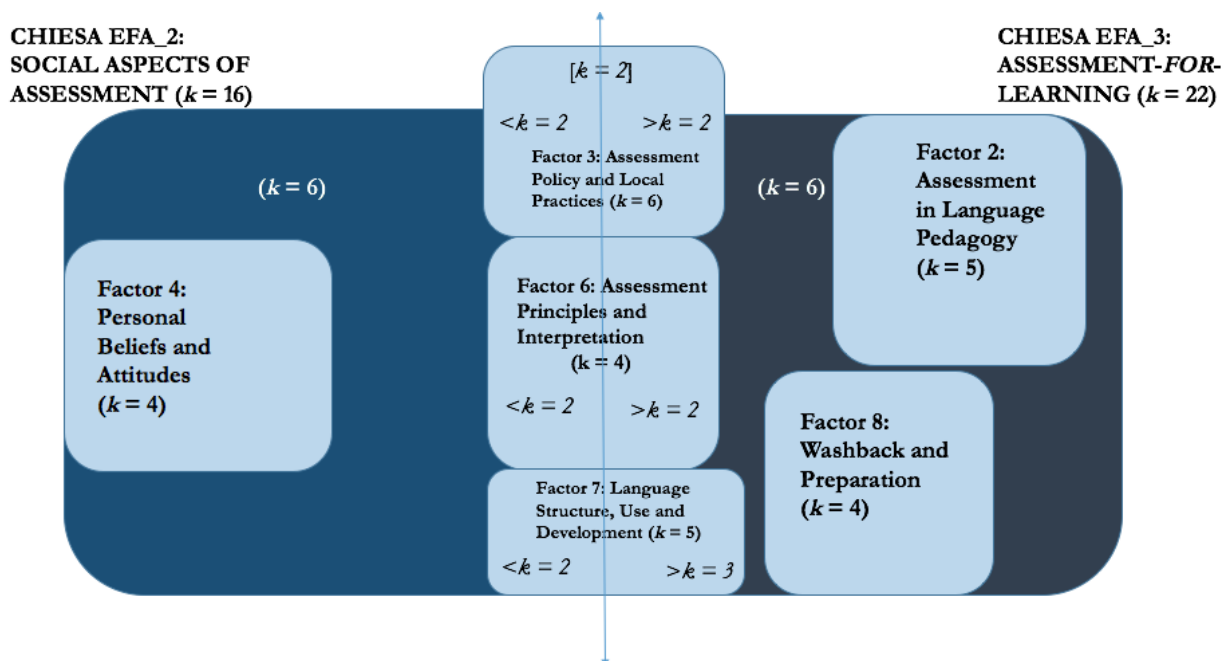


Figure 19 Chiesa, Factor 2 (blue): Social Aspects of Language Assessment; and, Chiesa, Factor 3 (black): Assessment-for-Learning

Note. The (<) symbol indicates that the number of items in K&H's factor also belong to Chiesa's Factor 2; the (>) symbol indicates that the number of items in K&H's factor also belong to Chiesa's Factor 3. The [] brackets indicate that the number of items belong in K&H's factor but neither Chiesa's Factor 2 or 3.

Six items from Chiesa's second factor were not present in Kremmel and Harding's EFA: Item 14 – how to communicate assessment results and decisions to students or parents; Item 18 – how to find information to help in interpreting results; Item 27 – how pass-fail marks or cut-scores are set; Item 32 – the philosophy behind the design of a relevant language assessment; Item 33 – the impact language assessments can have on society; and Item 40 – different forms of alternative assessments (e.g., portfolio assessment). Additionally, all four items from Harding and Kremmel's Factor 4 were a part of Chiesa's second factor. Three items [Item 35 – the assessment traditions in a local context; Item X – how to determine if a language assessment aligns with a local system of accreditation; Item X – how assessments can be used to enforce social policies (e.g., immigration, citizenship)] were a part of Kremmel and Harding's third factor, but neither in

Chiesa's second or third factor. Additionally, Item 26 – *how social values can influence language assessment design and use*, was a part of Kremmel and Harding's seventh factor, but not in either Chiesa's factor 2 or 3. These results suggest that the relationship between social aspects and assessment are somewhat difficult to capture in a survey for the Uzbekistan EFL teacher population. Thus, the underlying dimensions between the measured variables (e.g., the scores given by the participants on survey items that correspond to social aspects) and latent constructs (i.e., underlying variables) are not as clear-cut as seen above in the extracted factors of Kremmel/Harding that correspond to Chiesa's first extracted factor.

Fifteen/Sixteen items of 22 survey items overlap among Chiesa's Factor 3 – Assessment-for-Learning and Kremmel and Harding's Factor 2 – Assessment in Language Pedagogy; Factor 3 – Assessment Policy and Local Practice; Factor 6 – Assessment Principles and Interpretation; Factor 7 – Language Structure, Use, and Development, and Factor 8 – Washback and Preparation. (See Figure 18 above for a visual representation.) (One item in the original survey on self- and peer assessment was split into two items for Kremmel and Harding's EFA.) However, six items (2, 8, 9, 11, 13, and 16) are a part of Chiesa's Factor 3 and not Kremmel and Harding's EFA: Item 2 – *how to use assessments to evaluate progress in language learning*; Item 8 – *how to interpret measurement error*; Item 9 – *how to interpret what a particular score says about an individual's language ability*; Item 11 – *how to determine if the content of a language assessment is culturally appropriate*; Item 13 – *how to communicate assessment results and decisions to teachers*; and, Item 16 – *how to recognize when an assessment is being used inappropriately*. Additionally, Kremmel and Harding's Factors 2 and 8 are completely submersed into Chiesa's Factor Three. These results suggest that there is a relationship between Chiesa's Factor 3 and Kremmel and Harding's extracted factors (2, 3, 6, 7 and 8). Also, Kremmel

and Harding's extracted factors show the nuanced nature of assessment literacy for classroom language teachers, particularly in the separation of Assessment in Language Pedagogy from Washback and Preparation. These extracted factors (and the items within them) reflect the different assessment practices classroom language teachers do inside and outside the classroom.

6.3.8 EFA Conclusion

In this section I addressed the research question, "What is the underlying factor structure of the Language Assessment Literacy Survey?" To answer this question, I conducted an exploratory factor analysis (EFA) with principal axis factoring, used an oblique rotation method, and achieved a factor structure that was identified as having good fit indices and strong reliability. The EFA of the LAL Survey uncovered the nuanced nature of the Uzbekistan EFL language teacher's LAL literacy, by presenting three categories (Technical Skills of Assessment, Social Aspects of Assessment, and Assessment-*for*-Learning) instead of the eight intended theoretical constructs the survey was initially seeking to measure: Knowledge of Theory, Language Pedagogy, Scores and Decision Making, Sociocultural Values, Technical Skills, Principles and Concepts, Local Practices, and Personal Beliefs/Attitudes.

Then, with a comparison of the results of my EFA with Kremmel and Harding's, I show that the Uzbekistan EFL teachers think similarly with other language assessment stakeholders around the world. (The exact number of participants from Kremmel and Harding's study is not known, but I do know that they had over 1,000 participants.) Within each extracted factor from Chiesa, Kremmel and Harding extracted factors that presented similar ideas, and thus, a similar relationship can be identified. Chiesa's Factor 1 of Technical Skills of Assessment corresponded highly to Kremmel and Harding's Factors 1 – Developing and Administering Language Assessments; Factor 5 – Statistical/Research Method; and, Factor 9 – Scoring and Rating.

Chiesa's Factor 2 – Social Aspects of Language Assessment did not correspond as high, but was similar to Kremmel and Harding's Factor 3 – Assessment Policy and Local Practices; Factor 4 – Personal Beliefs and Attitudes; and Factor 6 – Assessment Principles and Interpretation. Chiesa's Factor 3 – Assessment-for-Learning corresponded highly with Kremmel and Harding's Factor 2 – Assessment in Language Pedagogy; Factor 7 – Language Structure, Use, and Development; and, Factor 8 – Washback and Preparation. Overall, these results build upon the existing literature in the language testing field that examines how we understand and define the assessment literacy/language assessment literacy construct, and how we identify, evaluate, and respond to varying user needs, particularly language teachers.

6.4 Research Question 1 Conclusion

A valid survey is meant to accurately address what is being measured – the phenomenon under investigation. In this case, Kremmel and Harding's Language Assessment Literacy Survey was created to empirically validate different components that compose the language assessment literacy construct for different stakeholders in language assessment (Taylor, 2013). The study addressed the stakeholder population of classroom language teachers, particularly Uzbekistan EFL teachers. Thus, this chapter asked the following research question – To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy? As exemplified throughout, there were layers of ambiguity that suggest the survey in its initial form (i.e., the version used with the participating Uzbekistan EFL teachers) is somewhat of a valid measure for classroom language teachers. Additionally, the participants responded positively to each item in the survey, in which the measures of dispersion and central tendency emphasize that there is some variability

in the data set – within and between the items – but not a lot, which provided a lack of actionable information about the teachers’ language assessment literacy.

Through an external analysis of the language assessment literacy literature, I created my own operational definitions of the constructs that were used to code Kremmel and Harding’s survey. The items in the survey appeared a bit vague for me and the two other coders. A few items were categorized as two different subconstructs because the initial categories did not create clear boundaries among them (e.g., Language Pedagogy and Scores and Decision Making). I then compared the items I identified as problematic (i.e., ambiguous) with Kremmel and Harding’s factor analysis results. Most of the items I identified as controversial, in fact, double-loaded onto their EFA’s extracted factors and were thus removed. Harding and Kremmel’s factor analysis of 50 items – instead of their original survey of 72 items – removed most of the troublesome items I identified.

I conducted my own Exploratory Factor Analysis (EFA) to explore the underlying factor structure of the survey, to see how the Uzbekistan EFL classroom teachers responded. I then compared my EFA results with Kremmel and Harding’s. The EFA uncovered three categories: Technical Skills of Assessment; Social Aspects of Language Assessment; and, Assessment-*for*-Learning. Subsequently, when I compared my factor analysis results with Kremmel and Harding’s, I was able to identify a similar relationship between the two, but not completely parallel results. This relationship shows that the initial survey I used with the participating Uzbekistan EFL teachers generally measured similar subconstructs to what Kremmel and Harding’s study found.

6.5 Discussion and Limitations

As explained in Chapter 2, surveys are perceived as easy to construct (Wolf, Joy, Smith, & Fu, 2016). After careful consideration of the values that underlie surveys and the amount of time necessary for item construction and validation, one may be able to see the inherent complexity of trying to design items that address the specific construct(s) being measured (e.g., Language Assessment Literacy). First, I discuss whether the results match Taylor's (2013) overall premise. Second, I discuss the limitations of the study. Finally, I discuss survey recommendations for the use of the Language Assessment Literacy Survey with underrepresented populations of classroom language teachers around the world.

6.5.1 Comparing to Different Stakeholders

Taylor's (2013) main premise is that different stakeholders in language assessment will think certain skills/knowledge about language assessment are more necessary than others. She conceptualized these differences in a visual representation (see Chapter 2, for Taylor's Levels of AL/LAL differentiated according to stakeholder constituency).

As explained in Chapters 2 and 4, Kremmel and Harding's Language Assessment Literacy survey was created to not only empirically validate Taylor's profile descriptors, but also identify the assessment training needs of these different stakeholder groups. Kremmel and Harding surveyed multiple stakeholders ($N > 1000$), including groups of people that were not originally mentioned by Taylor (2013) (e.g., parents of test takers). Ultimately, they reported surveying seven different stakeholder populations with the same data collection instrument: language teachers, language test developers, language testing researchers, applied linguists, policy makers (educational politics, university admissions, immigration), test takers, and parents of test takers.

There are pros and cons to surveying multiple stakeholders with the same survey. For instance, surveys that examine multiple stakeholders are useful in capturing divergent experiences, perspectives, and interests. These types of surveys can strengthen relationships and foster a culture of collaboration amongst various stakeholder groups. In other words, if researchers are able to identify the differences among stakeholder constituencies, we can learn how to better support converging and sometimes diametrically opposed relationships. A con to examining multiple stakeholders with one type of survey can be about stakeholder accessibility. With populations that are difficult to access (e.g., parents of test takers), a researcher might need the support of other stakeholder populations (e.g., test takers).

Participants from Kremmel and Harding's Language Assessment Literacy survey were asked to select from a list of a group/profession at the onset of the survey. They were also told to think about their involvement in this group/profession as they answered each closed-response item. Although Kremmel and Harding have not yet reported their results about whether the survey was able to measure differences among stakeholder groups, the results of the survey from Uzbekistan EFL teachers can contribute to the discussion. Responses from this group of teachers – a homogenous population of Uzbekistan EFL teachers – support Taylor's (2013) idea that classroom teachers believe they should have more knowledge/skills (related to assessment) with language pedagogy. The results also suggest that instead of the theory informing the survey, the survey informs theory, and how we conceptualize the AL/LAL construct should be extended to classroom teachers with more emphasis on Language Pedagogy and Assessment-*for*-Learning aspects. Specifically, participants of the study strongly endorsed items that addressed language, the relationship between teaching and assessing, and providing feedback. All of these aspects are important (and viewed as important) to a classroom language teacher's profession. The

participating Uzbekistan EFL teachers did not endorse as strongly items that addressed their personal beliefs/attitudes, using statistics, or communicating results to students and parents. These items could be endorsed more strongly by a different stakeholder population.

By looking closely at the degree to which the stakeholder population values certain components of language assessment literacy, researchers can gauge if certain stakeholder populations view assessment skills/knowledge differently or similarly to classroom teachers. I cannot say if all of the spider webs, in fact, work as they were intended to because my data does not extend to other assessment stakeholder populations. However, the overall premise is that my results show that Uzbekistan EFL teachers value and place more emphasis on concepts that are closer to their experiences (e.g., assessment in language pedagogy) than farther away (e.g., using statistics).

6.6 Using the LAL Survey for Research and Practice

Kremmel and Harding (*forthcoming*), Scarino (2013), and others who examine the language assessment literacy of classroom language teachers are one step closer to understanding and defining the AL/LAL construct. Taylor (2013, p. 406) questioned if we should understand and define assessment literacy for classroom teachers in terms of a more integrated concept, such as in Xu and Brown's (2016) *Teacher Assessment Literacy in Practice* (see Chapter 2). More work, I believe, should continue in this line of research with language teachers from all over the world. Using survey research to investigate all populations (e.g., underrepresented) of EFL teachers' language assessment literacy is a worthwhile and interesting endeavor.

Using Kremmel and Harding's Language Assessment Literacy survey in Uzbekistan, I was able to identify how an underrepresented EFL classroom language teachers' cognitions around assessment literacy (i.e., the assessment needs of teachers) are similar to, or different

from others stakeholders in Kremmel and Harding's (*forthcoming*) study. Thus, using the same data collection tool as Kremmel and Harding, I was able to strengthen the conceptions around the assessment literacy construct for classroom language teachers. From the study, I propose we emphasize *Assessment-for-Learning* when one wants to examine classroom teachers' assessment literacy. More specifically, I found that classroom teachers care more about the skills/knowledge that are closer to their everyday assessment/testing practices, supporting Taylor's (2013) main premise that different stakeholders in language assessment will think certain skills/knowledge about language assessment are more necessary than others. Kremmel and Harding's results have informed my study to divide *Assessment-for-Learning* into *Assessment in Pedagogy and Washback* and *Preparation*. These latter two constructs were reported highly by Uzbekistan EFL teachers, which could be identified as potentially important areas for all classroom language teachers around the world.

To use the Language Assessment Literacy survey with other underrepresented populations, for research and practical purposes, I recommend adapting the survey to each cultural context in which it will be used. First and foremost, one should pay attention to the purpose of the survey. Kremmel and Harding's Language Assessment Literacy Survey asks participants to provide responses to the following two inquiries: (1) How knowledgeable do people in your chosen profession need to be about each aspect of language assessment; and (2) How skilled do people in your chosen profession need to be about each aspect of language assessment? The information obtained from these two questions informs us about what classroom language teachers think or believe is important, and does not necessarily mean that they have the declarative and procedural knowledge themselves. A language teacher educator

could then adapt what the language teachers think or believe is necessary for them, and use that content to support language teacher development in that local context.

The second recommendation comes from Lietz (2008), who explains that when crafting items, survey researchers should identify the construct they are trying to target, and then be clear and concise while also providing the respondents enough options in their response so they have variability in the options. As explained in Section 6.2.2.2, there are a number of items in the survey that are ambiguous and vague. The items identified as ambiguous could be rewritten and more specific to the context in which they will be used. For example, Item 17 – *How to prepare learners to take language assessments*, is vague because it does not emphasize if the teachers' role in this case is to prepare learners to take large-scale standardized tests (such as the Test of English as a Foreign Language [TOEFL] or the International English Language Testing System [IELTS]), or the students' midterm/final/classroom-based exams. Perhaps, if a future researcher or teacher educator wants to examine (or educate) the language assessment literacy of understudied populations of EFL teachers and their LAL with standardized tests, then the item could be, “Item 17 – How to prepare learners to take TOEFL/IELTS;” or, if a researcher is only interested specifically in examining Assessment-*for*-Learning then the person would rewrite the item as such: “Item 17 – Prepare learners for classroom examinations (e.g., midterm/finals)”. Additionally, there could be a value judgment on the part of the researcher if the item is written as follows: “Item 17 – Taking class time to prepare learners for classroom examinations (e.g., midterm/finals)”.

The third recommendation is to provide the participants who take the survey with language support. First, a glossary of terms at the end of the survey to avoid language and content issues would be supportive. Also, I believe it would be helpful to bring multiple,

professionally translated, and agreed upon (by assessment specialists and professional translators) language versions of the survey to the research/teacher training site (e.g., English and local contexts' language[s]). Finally, one could assist the participants if they provide a “not sure/don't know what this means” category, which could inform future teacher educators of areas in language assessment that the population might benefit from exploring.

Kremmel and Harding's Language Assessment Literacy survey shows evidence of validity. In its current state, it can support future research in language assessment literacy, and it can provide support as a data collection tool to be used by teacher educators with new or different populations. Warwick and Lininger (1975) explain, “Every method of data collection, including the survey, is only an approximation to knowledge. Each provides a different glimpse of reality, and all have limitations when used alone” (pp. 5-6). Although there are limitations, I recommend continued use of the Language Assessment Literacy survey research in conjunction with other quantitative and qualitative research methods to understand classroom language teachers' assessment literacy.

7 RESEARCH QUESTION TWO RESULTS AND DISCUSSION

I present the findings of the qualitative analyses, which were conducted in order to address the following research question: How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students? To address this inquiry, I ask the following three subquestions: A. What do Uzbekistan EFL university teachers report they do to assess their students (i.e., assessment tasks and scoring procedures)? B. What are their cognitions surrounding those assessment practices? C. How do their cognitions about assessment shape how they assign scores to their students' work? These results include a discussion of both formal assessments (i.e., those that have points or grades attached to them) and informal assessments (e.g., comprehension checks), but most data sources will exemplify formal assessments.

The data sources came from focus group and/or semi-structured interviews. A focus group data source is identified by three conventions: (1) pseudonym of the teacher (see Section 4.3.1.1), (2) the teacher's number in the focus group, and (3) the number of the focus group. Here is an example of the prose conventions employed in the report: 'According to Dana in Focus Group 5, "the chance of cheating lowers when more test papers are created" (Teacher 5_FG5).' Thus, Dana (a pseudonym) is identified as the fifth teacher from the fifth focus group. There are two citing conventions to denote a semi-structured interview: (1) pseudonym of the teacher (see Section 4.3.2.2), and (2) the semi-structured interview number. Here is a second example of prose conventions: 'Ulugbek, who is an ESP teacher from Djizzak, said that "in many institutions, continuous assessment [is] forty [points out of 100 total possible points], but [each of our classes in my institution only assigns] thirty [points]" (Ulugbek_3). Ulugbek is the teacher's pseudonym and he was the third participant to be interviewed.

There were fifty-three participants in the focus groups, and twelve teachers were selected from those groups to take part in the semi-structured interviews. Sometimes, the same teacher (with same pseudonym) will be quoted from the focus group and semi-structured interviews, however, the naming convention will designate which data source it came from. The quotations of the participants were either (1) kept intact and not edited for spelling or grammatical infelicities, or (2) edited/paraphrased to make reading more comprehensible. When I edited quotations for clarity, I used brackets [] to denote my paraphrasing. Because the fifty-three teachers' English proficiencies varied greatly, some of the teachers' direct quotations had to be paraphrased in their entirety.

Throughout the five focus groups and twelve one-on-one semi-structured interviews, Uzbek teachers reported two broad areas of assessment: *Assessment-for-Learning* (i.e., Pre/During Learning) and *Assessment-of-Learning* (i.e., Post Learning). To check if each coded category could stand on its own, I used 15% of the transcriptions and conducted the card-sorting technique (as cited in Nunan & Bailey, 2009) with one other coder. As explained in Chapter 4, I placed individual statements from each category onto three-by-five index cards and placed them randomly in a pile. After I explained the definitions of each category to two additional coders, I had them distribute all the cards into the respective categories. Then, intercoder agreement was calculated. The intercoder reliability for the following categories [i.e., *Assessment-for-Learning* (Pre/During Learning) and *Assessment-of-Learning* (Post Learning)] was .98. I now discuss each reported area and provide relevant examples from the focus groups and interview transcriptions.

7.1 *Assessment-for-Learning* (Pre/During Learning)

The first category participants reported is *Assessment-for-Learning* (Pre/During Learning). This category is about assessments, both formal and informal, which provide information for teachers

and students *to identify the next steps for learning*. The teachers described doing two things: (1) continuous assessment and (2) placement/diagnostic assessments (see Table 30).

Table 30 Assessment-for-Learning (Pre/During Learning)

General Category	Subcategory	Number of Teachers Reporting
Continuous Assessment	Presentations	30 (56%)
	Assignments	30 (56%)
	Participation	29 (55%)
Placement and Diagnostic Assessments		24 (45%)

Note. 53 participants.

7.1.1 Continuous Assessment

Participants reported most of their assessment duties revolve around *continuous assessment* – the evaluation of students on a day-to-day basis, which includes presentations, assignments, and participation scores. All teachers are required to report to their institutions one continuous assessment score at the end of the semester, as a part of students’ total scores (i.e., a combination of continuous assessment, midterm, and final).

Thirty participants reported following the PRESETT (Pre-Service English Teaching and Training) curriculum for English Language and Literature – English Teaching majors. This curriculum was created by the British Council in 2013 and has been required by the Ministry of Higher Education for all teachers of future English teachers. The *PRESETT* Curriculum is thus designed to prepare future teachers of English to be competent in using the English language and in employing effective methods of language teaching. Courses include English language skills (e.g., listening/speaking, reading/writing, and vocabulary) and Teaching English as a Foreign Language (TEFL) (e.g., approaches to language teaching, language learning, classroom observation, and language testing/assessment.) The thirty participants reported their assessment practices are tied directly to this curriculum because it provides assessment specifications for continuous, midterm, and final assessment measures.

Twenty-three participating Uzbekistan EFL teachers reported they are English for Specific Purposes (ESP) teachers. These teachers follow an ESP curriculum, which is different at each institution. Similar to the PRESETT teachers, ESP teachers sometimes diverge from the set curriculum when they assign a continuous assessment score. Ulugbek, who is an ESP teacher from Djizzak, said that “in many institutions [for ESP curriculums], continuous assessment [is] forty [points out of 100 total possible points], but [each of our classes in my institution only assigns] thirty [points]” (Ulugbek_3). According to the nature of the courses that the Uzbek language teachers teach, and whether teachers are a part of the PRESETT curriculum or an ESP curriculum, there may be differences in the total possible points assigned for the continuous assessment score.

7.1.1.1 Continuous Assessment Cognitions

Some participants reported they find doing continuous assessment to be a valuable, worthwhile, and necessary endeavor, while others reported that they do not understand the purpose of continuous assessment and have a negative attitude toward it.

Ajva, Albina, and Leila, who have been teaching EFL for eight to ten years at a public university, held positive beliefs/opinions toward continuous assessment. They reported that they use the results from continuous assessments to make decisions about how to progress in their courses, and how to better inform their students on how to improve their language skills. During their discussion in Focus Group One, they explained how they interpreted (and employed) continuous assessment to all other teachers in the focus group:

Albina: So we have [continuous assessment], it means we assess them every lesson, each lesson. For example, how do they participate in each lesson, how do they do their home task, assignments ...

Leila: So do you do it just according to the lesson plan? And lesson plan is usually what they made according to the curriculum, according to fixed book, fixed literature?

Albina: No, for example during the lesson we do different tasks, different activities, and at the end [of the day], [we give them a score]. [The score for the day, plays a part in their final] continuous assessment [score]. We score them, but also help them [with their language skills].

Leila: Continuous assessment.

Ajva: Yes. We have only one final [score for continuous assessment] the end of each term [...] Continuous assessment helps us [teach] and students [learn]...

Ajva's comment that continuous assessment "helps us and students" emphasizes that the outcomes are meant to assist students in their language learning. During the discussion of continuous assessment in Focus Group One, all participants nodded their heads in agreement with Ajva, Albina, and Leila. They believe it is very important for teachers to know how to do continuous assessment, and they all had favorable/positive attitudes toward it.

However, unlike the participants in Focus Group One, some did not view continuous assessment as favorably. Those who did not approve were primarily over the age of 50. These language teachers had previously been teachers of Russian as a foreign language while the USSR was in control of Uzbekistan's education system. At the beginning of independence in 1991, they were told they needed to become teachers of English. Shaholo from Focus Group Four interjected the following into the group discussion when I asked the participants about how teachers' total scores of continuous assessment were broken down:

May I talk about my time when I was teaching in the time of the USSR? There was no continuous assessment. Because we had only two, three marks. [The marks were tests.]

Either you came to class or you did not, and that was not [scored]. But nevertheless the level of the knowledge was much more back then. Now we have different criteria, subcriteria, what the students should know during the lesson. We divide everything to the little, little details. Too much. All the items the students should acquire, let's say. But still, we pay a lot of attention. We all the time, we tried to modify this percentage, this points for continuous assessment. If you were not trained by British Council, who made PRESETT, you do not know exactly [how to score] [for continuous assessment] (Teacher 10_FG4).

Shaholo expressed her opinion about continuous assessment. Based on her experience during the USSR time in Uzbekistan, she believes the practice of continuous assessment is too complicated a procedure, and that it ruins the flow of teaching and learning language. During this time, language teachers focused on disseminating knowledge to students so they could help them pass tests, because passing tests was seen as the ultimate marker of achievement. Therefore, the new teaching practice of assessing students daily has been a challenge for some EFL teachers who were not accustomed to this practice. Not only does Shaholo have a mixed belief with respect to continuous assessment's effectiveness, but she has a negative attitude toward it. Shaholo's comments about the British Council suggest that she might have had a more positive/favorable outlook on continuous assessment if her university had provided her the opportunity to be trained. Her cognitions and feelings seem to be shaped by her university context and her past/current teaching experiences within two different political climates.

7.1.1.2 How Students are Scored on Continuous Assessment

In general, the younger generation of Uzbekistan EFL teachers find that doing continuous assessment is important and facilitates their students' progress in their English language ability.

The older generation of participating Uzbekistan EFL teachers, who taught languages during USSR's time in Uzbekistan, find that continuous assessment, in general, is a "waste of time" for students who need to solely concentrate on the knowledge and skills that the teachers teach them. Although the participating teachers reported varying cognitions and emotions about continuous assessment, teachers do not have a choice about whether to provide students with a continuous assessment score. The language teachers reported similar procedures for how they kept track of students' progress and eventually provided them with a score.

Kamila from Navoi, who has been teaching English for approximately fifteen years, uses the PRESETT curriculum at her university. Operating within this curriculum, she describes how she conducts and scores continuous assessment:

I have [one-on-one] conferences with them. [Also], I observe them [during the class] and I try to record everything [informally]. I keep my recording like something like diary maybe [...] I do not always assess students [in a] formal way [for continuous assessment]. Sometimes informal, [which is] my own feeling about the student"

(Kamila_8).

Kamila emphasized that she takes time to talk to each of her students. When she reported that she "records them" it does not mean that she audio records each of her conversations; instead, as she learns about their strengths and weaknesses, she writes her impressions in a diary to better inform what she did in the classroom. Kamila believes that continuous assessment helps inform her teaching practices. Kamila's discussion above explains that she provided a score based on her *feelings* about how the student did *overall* throughout the semester, and that she tried her best to keep track of their progress in a diary-like format.

Guldasta from Focus Group Two, is an ESP teacher at a private university in Tashkent, and also includes her feelings in students' continuous assessment scores. Her university does not have established criteria for evaluating, or defining, what continuous assessment is. She explained,

For continuous assessment, I just ask during the lesson the questions the [students] will answer, and their attendance, are they ready, have they understood – This is continuous assessment. [There is] no criteria that teachers use to give a [score]. Statistics, that is, [is just about giving scores]. But, anyway, I have some alternatives in order to understand the student. A number [or score] is not a representation of a students' abilities.

Both Kamila (PRESETT) and Guldasta's (ESP) experiences with continuous assessment illustrate that when teachers are required to provide continuous assessment scores, they often assign scores based on their intuition or general impressions, regardless of how they feel about the practice of continuous assessment. These teachers, and the rest of the participating Uzbekistan EFL teachers, appear to believe in their ability to understand each student's progress through presentations, assignments, and participation.

7.1.2 Presentations

Thirty Uzbekistan EFL teachers reported asking students to do presentations as part of their continuous assessment scores. Of these teachers, nineteen ESP teachers and eleven PRESETT teachers reported using presentations. Tahmina, who is a chemical engineer ESP teacher in her second year of teaching English, emphasized that “[As ESP teachers in a science school, our university provides us much material that is related to chemistry. However, they also provide us materials for] tourism, internet [and] food, for example. [We] use these materials for presentations. [We make] students do presentations. Lots of presentations” (Tahmina_FG4).

7.1.2.1 *Cognitions about Presentations*

Some participants reported they find doing oral reports as a part of the continuous assessment score as a valuable activity that helps students learn specific language forms, informs teachers about their students' language issues, increases motivation for learning language, and is a great way to support the teachers' lesson by connecting their classes to the real world.

More ESP teachers reported using presentations as part of a continuous assessment score because they believe that they are one of the most helpful techniques to learn specific language forms. Aziza, who is in her first year of teaching English at an engineering university in Karakalpakistan, explained that she and her colleagues "... made students [do presentations]... although I can state it also, it was a bit difficult, but it was very helpful for their language" (Aziza_7). Aziza believed that speaking at the front of the room is an effective way for students to learn language.

When I asked her what she meant by "it was a bit difficult" (Aziza_7), Aziza emphasized that as a teacher, she found it hard to convince students that presentations are helpful for their language learning. Furthermore, I asked Aziza to clarify how she identifies her students' language concerns during presentations, and how she then informs them afterwards of issues or concerns that they should continue to work on. She iterated that "presentations help students practice speaking" (Aziza_7). I ended the conversation without pressuring her to be more specific about how she understood language skills to improve. Aziza's interpretation of the use of presentations emphasizes a belief that *the more practice, the better for language development*, but she seemed unable to clarify how she identified her students' errors and/or how she informed them to better assist in their language development. Aziza believes that presentations are an

effective way to learn language, but, she was unable to articulate how she used presentations to help students do so, which could have been due to her English proficiency level.

Along with Aziza, most participants did not explain how they obtained and used the information they gained from the presentations to assist in helping their students learn language, change their syllabus, or inform their teaching practices. Most participants emphasized that presentations are a “fun way to promote student motivation” (Aziza_7). There is an understanding among some participants that language skills will improve if more opportunities are provided to students to speak in a fun and exciting way that could be seen as motivating (e.g., through the use of in-class presentations). The Uzbekistan EFL teachers who discussed “fun” were all novice language teachers in their first couple of years teaching after having completed their university degrees.

In addition to language learning as a motivating factor, teachers believe that using presentations was a great way to connect the real world to their classes. Darisa in Focus Group 3, who is in her fourth year of teaching English, believed that presentations are a great way to connect the topic in the class to the news of the day.

Well, first of all I want to speak about lessons. [In] the beginning of our lessons, I ask about the current events and the news all over the world. [Students] will [say] political culture, econom[ic], and [about the environment], news maybe, and others. Maybe a visit of our president to other countries ... I [adapted] this technique from my university teacher. Then, [students are asked to use the internet to gain more information], yes. And then, [in each class] students in groups will do short presentations on the news of the day. They are fun and interesting. (Teacher 2_FG3).

Diesia's use of presentations was a way for her to connect what was going on in the world outside of her classroom and with what they were studying in class. She, like Aziza, found presentations to be a fun way for students to engage in the content of language lessons.

Using presentations is perceived by some Uzbekistan EFL teachers as a helpful way to provide formative assessment on students' language skills, and they believe that presentations are a good way to support teachers' lessons. However, these teachers seemed unable to explain fully and articulate *how* they identified their students' language issues, and how they provided information to their students to better support their English development. The participants who reported positively on using presentations were all beginning (novice) teachers of English, who recently graduated from university with a degree in teaching English.

7.1.2.2 How Students are Scored on Presentations

The participants who reported doing classroom presentations as part of their continuous assessment score tended to be teachers who were younger in age and held a general belief that the more language practice, the better. This understanding was the basis for how some teachers scored students. Nodira from Focus Group Three explained that in her class, "The students who talked more or who sounded [more fluent] got good [scores]" (Teacher 8_FG3). Not all participants were as forthcoming as Nodira was about how they scored students' oral reports.

When I asked the participants how they score students' presentations, most teachers reported that they did not state explicitly that they use a rubric. Sakina from Focus Group 4 told the group that she is "not too strict... [I] have a list of topics [that I think are] important such as grammar and voice, but [nothing] formal" (Teacher 11_FG4). Most teachers from focus Group 4 then murmured in agreement, "mmm" and "hmm mmm." The participating Uzbekistan EFL

teachers overwhelmingly reported they gave scores based on their overall interpretation of how the students did during their presentations.

Yayra, however, from Focus Group 5 explained that scoring presentations “would be more feelings based, but now teachers are aware of assessment rules, they have gone through many trainings on average” (Teacher 10_FG5). Yayra is an EFL teacher who has been to the U.S., U.K., and Australia. In addition, she has participated in numerous U.S. Department of State’s EFL training courses in language teaching methodology. She lives in Tashkent, the capital of Uzbekistan, where most (if not all) trainings on language assessment have happened and has had the opportunity to take part in many of the language teacher educator trainings. Yayra’s statement shows that trainings in language assessment are entering into Uzbekistan’s EFL teacher preparation and in-service teacher education courses. According to Yayra (Teacher 10_FG5), traditional classroom assessment practices that are experienced in the West are slowly taking on a larger role in these Uzbek teachers’ lives. Yayra explained that “[Language] teachers are more aware of specific things to [score]. More detail focused than general focused” (Teacher 10_FG5). Yayra was making a claim that the teachers are beginning to change their assessment practices when it comes to classroom presentations, in which the use of rubrics is beginning to take the places of teachers’ more subjective intuitions and/or feelings.

To understand more fully how different teachers across Uzbekistan (with respect to both geographical region and age) use and score presentations, I talked with Bibidana and Olmar from Focus Group 3. Bibidana had been teaching English for 33 years in Bukhara and Olmar for 29 years in Andijan. Bibidana told the group that she does not usually use presentations in her languages class because it “takes up too [much] time with no real benefit” (Teacher 1_FG3). The “no real” benefit Bibidana is referring to is language skills. She views presentations as a waste of

time because she believes that language skills are not being developed, and thus, she has a negative attitude toward presentations. When she does use a presentation activity to count as part of the continuous assessment score, she (and Olmar) score very similarly. Our exchange is below:

Interviewer: So, let me ask you two about how you mark [score] presentations or projects.

How do you do you evaluate it?

Bibidana: How do they work in cooperation

Interviewer: So the first criteria usually is how well they cooperate?

Bibidana: Yes.

Interviewer: Do we all agree with that?

Olmar: How do they cooperate in a group? How do they just.... How can I say, behave [in the group], maybe with the materials, maybe with just presenting, outputting.

Bibidana and Olmar's first reaction was that they scored presentations based on how all the students cooperated with each other. Examining cooperation as a criterion for presentations shows that these Uzbek EFL teachers place emphasis on group cohesiveness and interaction, more than on language ability or skills. (I will note that I am uncertain if they truly see cohesion/synergy as building language and that they were unable to articulate it so.) Bibidana and Olmar did not discuss further criteria for evaluating the presentations, but emphasized that how students worked with each other was an important component of how they viewed the students, and subsequently, provided them a score. Presentations were one form of continuous assessment that Uzbekistan EFL teachers use in their classes. The second type are different types of assignments.

7.1.3 Assignments

There are four types of assignments the participants reported they used as part of their continuous assessment score: (1) *internet assignments* – activities that asked students to use the internet, (2) *writing portfolios* – a collection of written materials over the course of a semester or year, (3) *peer assessments* – evaluating partners’ works, and (4) *self-assessments* – evaluating students’ own work.

7.1.3.1 Cognitions Around Assignments

Some participants reported they find doing internet assignments helpful because it informs them about their students’ abilities with real language use (e.g., can students read/navigate around English websites and/or speak/write about their reflections, thoughts, or ideas). The use of internet assignments was favored by Uzbekistan EFL teachers because the internet has been identified by the participants as important in students’ lives. Students use their phones, personal laptops, and/or desktop computers to access the internet daily. Nozliya from Syrdara commented that, “students [are] attached at the hip [to] the Internet, and we should [use it] more” (Nozliya_11). Aida from Focus Group Five also explained, “The internet is helpful for our teaching because of modern technology. Young people want to know English language [more] than Russian language because all modern technology, mobiles, computers, everything is in English. And internet, they want to read something, majority of information in English” (Teacher 3_FG5). An immediate response to Aida from Chinara in Focus Group Five was, “[I think the same way you do! Most Uzbek students believe the English language is beneficial for navigating the internet. So, I like to use the internet for my class assignments]” (Paraphrased for clarity, Teacher 4_FG5). Although some teachers commented on the internet as an affordance to learn

language and content, only one teacher, Eldman, discussed what he actually did with the assignments to improve language learning or teaching.

Eldman from Focus Group 5 taught English for 5 years in Tashkent, and as a novice teacher, he explained he uses internet assignments in his classes because he believes the internet increases student motivation and helps him measure students' knowledge about real language use. Eldman decided to deviate from the PRESETT curriculum because he noticed that students in the past enjoyed learning about topics that were not a part of the mandatory curriculum. He emphasized that "When I taught them the topic restaurants...I couldn't do the PRESETT curriculum's home task assignment that asked students to write about restaurants in Uzbekistan. Instead, my home task was asking students to find unusual restaurants all over the world, and write about it ... They could surf on the net. However, it was a bit difficult for them to explain why each chosen restaurant was unusual. But, they learned about unusual restaurants and it was a very interesting project for them" (Paraphrased for clarity, Teacher 1_FG5).

The students' interests in completing the internet assignment motivated Eldman to continue doing this assignment for subsequent courses. He also explained that his students "received comments about language from me" and that they had "great ideas that I share" (FG5). In other words, Eldman made an indirect reference to language ability when he commented, "[They could surf on the net. However, it was a bit difficult for them to explain why each chosen restaurant was unusual]" (Paraphrased for clarity, Teacher 1_FG5). He used these cursory results to inform his teaching practices, in which he discussed the grammar point of comparatives and superlatives. After teaching his students about the comparative and superlative forms, meanings, and uses, he had them revisit the same internet websites about the restaurants they previously compared. Then the students were asked to rewrite their compositions based on the new

understanding of comparatives and superlatives. Eldman used Internet assignments as an assessment-*for*-learning technique, because he was able to learn about his students' grammatical mistakes through an internet assignment that students found to be interesting and motivating. Finally, they had to revise their work based their newly acquired grammatical understanding. In addition to internet assignments, EFL teachers in Uzbekistan also used writing portfolios as a way to learn about their students' language ability.

7.1.3.2 Cognitions about Writing Portfolios

The second type of assignment for continuous assessment discussed by the participants is writing portfolios – a collection of written materials over the course of a semester or year. Three Focus Group 4 teachers (Zamifra, Zulnara, and Farrukh), who taught reading from an ESP curriculum, did not know what comprised a writing portfolio, or what its different purposes might be. Most participants who taught ESP had heard of writing portfolios but did not use them in their teaching practice. In contrast, the Uzbekistan EFL teachers who taught academic writing or introductory composition classes in the PRESETT curriculum used writing portfolios. These teachers, although required by the PRESETT curriculum to use portfolios as tools for learning with their students, found they were a very valuable assignment. They believed they could identify students' progress in writing, learn more about their students' interests, and identify students' critical thinking skills. None of the teachers who used writing portfolios as a regular part of their teaching expressed any negative emotions about them.

Three mid-career level teachers who taught in the PRESET curriculum discussed their ways of using writing portfolios. Abdulaziz from Focus Group 2 emphasized that “[the writing portfolios] should [have many] entries, [for example], separate entry one, entry two, [and a] critical review” (Teacher 6, FG2). Dilyia, a writing teacher, explained that for her class,

“[Students should submit their portfolios with at least three entries. One entry is considered their best writing during the semester, then they should select two more entries on given topics provided during the semester. Ten points are assigned for each one]” (Paraphrased for clarity, Teacher 4_ FG2). Abdulaziz’s classes writing portfolios are different from Dilyia’s because he asked students to submit a critical review of one of their writings. He was interested in how students’ thought about their writing, and whether they could explain how to improve their writing. Klara from Focus Group 1 used the portfolio as an assignment because it “showed me a lot about students’ minds. Why they chose writings. Why they chose that exact one. Interesting and makes me think about students’ [language level and interests]” (Klara_ FG1). The writing portfolios contributed to the teachers’ knowledge about their students’ writing abilities, and the participants reported they were also able to identify the students’ interests. In other words, teachers could see from the students’ choices where their interests reside. However, the Uzbekistan EFL teachers did not explain how they used the information about students’ interests, or what they gleaned about their students’ language levels, for informing their teaching practices. Two assignments the writing teachers utilized as part of the portfolio assignment were peer and self- assessments.

7.1.3.3 Cognitions around Peer Assessments

The writing teachers did not discuss their beliefs, opinions, or ideas about peer assessments. Instead, the participants explained that they are required to do use peer assessments. All teachers who commented on peer assessment taught writing from the PRESETT curriculum. No teacher who taught ESP discussed peer-assessment. Durдона in Focus Group 3 commented that in all her writing classes, students are asked “to evaluate their partner’s tasks” (Teacher 3_ FG3). Evaluating writing is based on a checklist supplied by each teacher.

A peer evaluation checklist was reported only to be used in writing classes. A peer-evaluation checklist is a sheet of paper provided to each student that contains criteria the writer is to have included/met to have produced a successful piece of writing. Svetlana explained that, in her course on English Writing for English language majors at the Bachelor level, a peer-evaluation checklist is used during the mid-course assessment. “Total score is 30 [points] for mid-course assessment. [Students’ writings are scored out of] 20 [points] for writing [and] 10 for [the] self-evaluation checklist or peer evaluation checklist. [The choice of which type of checklist to use] depends on semesters” (Svetlana_2). I then asked Svetlana if she uses the peer-assessment evaluation checklist to make decisions about what to do in her writing class. Svetlana replied that “I don’t use it [to inform my teaching]. [It is just used by students, for themselves]” (Svetlana_3). Svetlana’s comment about her non-use of the peer-evaluation checklist shows that this material is only used when teachers are required to do so (e.g., if one teaches writing).

Svetlana’s comments reveal that peer-assessment is a required writing portfolio component that teachers in the PRESETT curriculum must use. Svetlana emphasizes that what the students do in this peer-evaluation does not have an impact on her teaching practices or how she intends to inform her students about how they can improve their language skills.

7.1.3.4 Cognitions around Self Assessments

Self-reflection is a human capacity to engage in processes of introspection, and students in Uzbekistan are often asked to “think about the writing they do” (Teacher 5_Focus Group 2). Self-reflection tasks were not reported by teachers who work in an ESP curriculum; instead, two writing teachers, who work with students who are majoring in English language and literature, reported they use self-reflection in connection with their writing classes.

Self-reflection tasks were reported to be completed in two ways. First, Guldasta in Focus Group 2 explained that she uses a self-reflection checklist “which is for mid-course assessment and the total score is 30, two tasks, writing test 20 and self-evaluation checklist 10” (Teacher 8_Focus Group 2). Second, Kamila reported that she uses self-reflection in her writing class as a self-reflection log, in which, students write a “200-word reflection on their writing about words and grammar and if their grammar got longer and words more complicated. Finally, students need to say how they will improve as a writer” (Kamila_8). Both Guldasta and Kamila emphasize the importance of self-reflection tasks as part of their writing classes. Similar to the peer-assessment task above, the Uzbekistan EFL teachers who utilized this assignment did not explain what they did with the students’ work. I asked the teachers if they made logs or checklists about different topics that the students find interesting (or not) and whether they read through the peer-evaluation and self-assessment assignments. No teacher responded to my queries.

7.1.3.5 How Students are Scored on Assignments

Depending on the school, specific program, and whether they use PRESETT or are involved in an ESP curriculum, points are allocated differently for each type of assignment.

In general, the participating Uzbekistan EFL teachers reported they do not use rubrics to score assignments, but instead, focus on their *impressions* of students’ work. Their impressions come from comparing students’ completed assignments to others in the classes. Additionally, the students’ *speaking* ability plays a role in how scores are assigned. I asked Mohira, a novice teacher who has been teaching EFL for one year in Kashkadarya, about how EFL teachers at her school score assignments. An excerpt of our conversation is presented below:

Interviewer: ... what are some factors that influence a teacher's impressions on grading assignments?

Mohira: Mostly they refer to speaking abilities. It doesn't matter if even they're teaching writing. For example, if I am teaching writing, there are some students who have been in U.K., in U.S., and they come, but they don't have appropriate writing skills. They have spelling mistakes. We teachers don't have [appropriate] writing skills... So, I will provide a [score] with their speaking.

Mohira explains that students' speaking ability plays a large role in how their assignments are scored. Additionally, her comments show that she doubts her own English language skills. If she has a student whose productive (e.g., speaking/writing) skills are stronger than hers, then she will defer to them and provide them with a higher score. Thus, she tends to score students based on her emotions. The score she assigns seems grounded in general impressions compared across the students in her classes *and* between herself and the student.

One teacher, Madison, who has been teaching for 10 years at an ESP institution in Tashkent, reported that teachers at her university now rate students in a different way than previously. “[The teachers no longer score by their impressions.] There was a time when they did this, but now they use a more analytical approach to scoring. So, they score analytically *and* holistically” (Teacher 4_FG3). What should be noted is that the participants who discussed that the teachers no longer use impressions to score students are the teachers who live and teach in the more populous cities in Uzbekistan and at the schools that are considered the better-quality institutions.

The participants who reported using internet assignments and writing portfolios believe these assignments are helpful for both teachers and students. First, they are valuable because they

serve to inform teachers about students' abilities with language use. Additionally, teachers reported they make it possible to better identify students' progress in writing. In other words, they can learn more about students' interests and more easily identify students' critical thinking skills. No teacher who used internet assignments or writing portfolios expressed negative emotions about their use. The Uzbekistan EFL teachers tended to use a neutral tone when they discussed peer and self-assessments. No teacher reported they found these tasks effective or worthwhile, but they commented that they are required to use them because they are a part of the PRESETT curriculum. Scoring assignments used to be based solely on the teachers' general impressions. However, not all Uzbekistan EFL teachers use this approach. Some who have been trained in assessment (e.g., teachers in the capital, Tashkent) now will try to utilize a rubric or some other form of a more scientific approach to scoring. There was much discussion among the participants about one aspect of the continuous assessment score – participation.

7.1.4 Participation

All Uzbekistan EFL university teachers reported evaluating participation on a day-to-day basis.

7.1.4.1 Cognitions about Participation

Participation was reported by the participants to be a vital part of continuous assessment scoring because it helped Uzbek teachers “know which students took their language learning seriously” (Diora_4). Nilufar, from Focus Group 2, has been teaching English for 40 years, longer than any of the other participants in the study. Nilufar explained within the second Focus Group that “... most people in Uzbekistan, especially those that are older, and those people from other post-Soviet countries, don't believe in independent and autonomous learning... they believe that *real* learning can happen only under strict teacher supervision” (Teacher 11_FG2). Thus, she emphasized that teachers of the older generation view their role in the classroom, and in the lives

of the language students, as one of the most important contributors to successful language learning. Many teachers nodded their heads in agreement with Nodira's point about why participation is such an important concept in the Uzbekistan education system – *The belief that learning can only happen under the guidance of someone more knowledgeable*. If the student does not come to class, then he or she will not be learning. However, throughout all Focus Group discussions (1-5), there was considerable discussion about how to define what 'participation' meant, and then, how to appropriately score it. The PRESETT curriculum and each school that has an ESP focus do not have the same definition of participation. What we know for certain is that each school in Uzbekistan requires all teachers (of all subject/content areas) to provide students with a participation score.

7.1.4.2 How Students are Scored on Participation

Teachers reported two ways to assess participation: (1) interaction and engagement during the lesson and/or (2) attendance. The way a teacher viewed participation seemed to pivot on whether he or she belonged to a younger or an older generation of teachers.

Teachers who have been teaching from one to fifteen years reported participation as interaction and engagement during the lesson. Albina from Focus Group 1 discussed what participation meant to her: "Participation means that they should come to the lesson, it's not just attendance but they should come to the lesson and always they do their home assignments. [Then], and they will participate during the lesson, [in which] they will interact with each other, and answer the questions [I ask] and give their opinions. This [to me] is participation" (FG1). She defined participation as engagement in the teaching and learning of content during the class period, and that the student should be interacting in a cooperative manner with classmates. I attempted to delve further into this topic by asking her what strategies she used to remember how

each student participated. Did she “walk around and put a check mark by every student when they participate?” She replied, “No, just put some ticks. Always I can observe them and I know [how each] student works in my classes. Sometimes I have video recordings [to look at] but mostly [I will check them after] the lesson. [Finally,] I can put some scores on my [grade sheet]” (Albina_FG1). Most Uzbek university classes are electronically recorded, and teachers can view the recordings. Albina did not use the video recordings often, even though the recordings are readily available and she could look at them as often as she might need to determine if she was providing students with accurate participation scores.

Other Uzbekistan EFL teachers, who have been teaching English (and Russian) for fifteen to thirty years, define participation only as attendance, and do not consider a students’ level of engagement during class as a component of a participation score. Farhod, Ona, and Nargiza from Focus Group 3 tried to clarify with each other and the interviewer how they assess participation:

Interviewer: How do you assess ‘participation’?

Nargiza: How [students] participate in our classes [cross talk 00:19:50]

Farhod: Participation means how well they learned the material. [cross talk 00:19:52]

Ona: You mean attendance thing?

Farhod: Attendance? Or participation? You see attendance is just one [part of participation].

Ona: Hmmm. Okay. Only attendance [is considered participation] at my school.

Farhod: Uh, attendance [is] just one point, and homework is five points, self-study is four points [which makes a total of 10 points available for participation].

This brief interaction emphasized that some Uzbek language teachers do (and do not) differentiate between participation as either engagement or attendance. Thus, coming to class, for some teachers, is a form of participation, and for others, coming to class is only a part of the total participation score. Teachers from all five focus groups discussed whether participation should be considered engagement during class or only as attendance, or both. “It is easier to just give an attendance score as participation, because [it’s] objective” (Dana_FG5).

Svetlana, a PRESETT language teacher from Samarkand, who had been teaching English for 30 years reported differently from her colleagues, even from her colleagues of a similar age, on how she defined participation. She defined participation for all Uzbek language teachers as comprising four different components. First, they are asked to make sure each student completes all class work and homework; second, students should only speak English in the class; third, students should volunteer and ask questions often; and fourth, the students should work well (e.g., cooperatively) in pairs and groups (Svetlana_2).

7.1.5 Placement and Diagnostic Assessments

Participants reported that their Assessment-for-Learning (Pre/During Learning) responsibilities not only revolve around continuous assessments, but also *placement tests* – exams that determine which class one should be placed in, and *diagnostic assessments* – the evaluation of students’ strengths and weaknesses based on the goals and objectives of a course of study. Depending on the university, a student needs to enter with a certain level of English based on the Common European Framework of Reference (CEFR). Shaknoza emphasized the importance of knowing the CEFR because “understanding that Framework is considered by EFL teachers in Uzbekistan as the common language amongst all language teachers” (Teacher 9_FG1). Shaknoza provided an example of how the CEFR is discussed among Uzbekistan EFL teachers, “Oh, this school

accepts students at an A2 level, but this student is at a B1 level...” (Teacher 9_FG1). The CEFR is identified by administrators, language teachers, and parents as the guiding Framework to identify language levels, progress, and achievements. Once students are admitted into a university with a certain CEFR level (scored on their national state testing center exam), they will be informally tested again with a spoken one-on-one interview placement test. Then in each class, teachers are asked to perform a diagnostic assessment to see whether students know the different aspects that the course is designed to teach them. Many Uzbekistan EFL teachers take part in these university language placement tests, and everyone is expected to conduct a diagnostic test.

7.1.5.1 Placement and Diagnostic Assessment Cognitions

Participants believe that placement and diagnostic tests are important information-gathering tools. Teachers use such tools in order to understand students’ language levels and identify the specific language skills in which students are both strong and weak. Umida commented that these types of tests “could inform us of how we can alter our teaching plans, syllabus, and how much we can talk to our students in English” (Teacher 7_FG2).

However, some teachers place little faith in the validity of their school’s placement tests, and prefer to see proficiency scores of publicly-recognized standardized TOEFL or IELTS examinations to determine course placement. Sometimes, students will not need to take a placement test at the Uzbekistan university because, as Ona from Focus Group 3 mentioned, “Awesome students will have [a score of] 6.5 on [IELTS]; [and some have a score of] 7, [which is way] over the mark” (Teacher 6_FG3). In other words, even though the English proficiency tests of IELTS and TOEFL are sometimes used as placement tests at some universities, the relationship between these tests and the universities’ curriculum is unknown and has not been

officially established by the Ministry of Higher Education. There is a commonly held belief among these teachers that large-scale standardized tests take precedent over their own exams that they administer and score. Two of the forty-one participants who are ESP teachers reported directly that their schools use IELTS and/or TOEFL scores in lieu of local tests developed by the Uzbek state testing center. These two teachers also informed Focus Group Three that they require all of their students to take the TOEFL and/or IELTS test. Three Uzbek teachers from across all focus groups emphasized that tests generated by the “British Council and ETS are always preferred” (Durdona_FG3).

7.1.5.2 How Students are Scored on Placement/Diagnostic Tests

Many teachers take part in placement tests at their universities and then use the resulting diagnostic measurements in their classes. I talked with Focus Group 5 about the general process, and Nozilya emphasized that these placement tests are not scored with a number, but instead, students are placed into groups of high, medium, and low:

Interviewer: When students come to your University, they have a certain score.

Yayra Yes.

Interviewer: And then when they come to your class, you regroup them.

Yara: Mm-hmm (affirmative).

Interviewer: How do you regroup them?

Yayra: They are regrouped.

Nozima: No, in ... yes.

Mukaddas: In our case ...

Nozilya: [At] our University, we [ask] first year students [who] have just entered [to take a spoken language test]. So all [language] teachers [form a] commission. And

one by one we have a conversation with [each student]. According to their [speaking] level, we separate into groups, high [and] low, high [and] low, high [and] low. Then, teachers [conduct] another test in their [respective] classes [to make sure the students were placed appropriately].

This exchange showed that students are being placed in classes based on their proficiency with the spoken language. Their speaking level will also determine the class they will join for writing, listening, reading, and grammar. As a follow-up, I asked Mukaddas why other language skills are not assessed on the placement test. She explained that “at our school, we do not have time [to test] or score all different skills” (Teacher 7_FG5).

The Uzbek teachers reported they conducted diagnostic assessments in their classes once the school year starts, and some reported that they base their lesson plans on the outcomes. Umida from Karakalpakistan explained, “[individually, I try to plan my first activities according to the diagnostic test I give at the beginning of my course. For example, when I give the test, I then analyze the results and think about the students’ level ...I also think about my knowledge of each student, and some gaps the students might have]” (Paraphrased for clarity, Umida_5). Umida explained that she uses the information gained from the diagnostic assessment to adjust her teaching practices. However, she does not explain directly how she changes the activities. Leila reported that many of the language teachers at her university “do diagnostics, but they don’t know what to do after that, how to change, or see their gaps” (Leila_5). Thus, many of Leila’s teacher-colleagues collect information about their students but have trouble implementing changes or reorganizing their classes based on diagnostic assessment results.

7.1.6 Conclusion – Assessment-for-Learning (Pre/During Learning)

The Uzbekistan EFL teachers reported Assessment-for-Learning (Pre/During Learning) as one of the measurement strategies they use at their institutions to assess their students. The information the language teachers identify from these measurements is intended to assist students to progress in their language learning. The two categories reported are (1) *continuous assessment* – the evaluation of students on a day-to-day basis that includes presentations, assignments, and participation scores, and (2) *placement and diagnostic tests* – assessments that identify students' language level in relation to different courses, and which also investigate students' strengths and weaknesses.

The Uzbekistan EFL teachers' cognitions around these assessment practices vary. Some reported they find doing continuous assessment to be a valuable, worthwhile, and necessary endeavor, while others do not seem to understand the purpose of continuous assessment and/or tend to have a negative attitude toward it.

Additionally, the participating teachers reported that presentations and assignments serve at least four useful purposes. They help students learn specific language forms, inform teachers about their students' language issues, increase motivation for learning language, and support the teachers' lessons by connecting their classes to the real world. All participating Uzbekistan EFL teachers believe that participation is one of the most valuable assessment-for-learning measures.

There are several identified factors that seem to influence Uzbekistan EFL teachers' cognitions about Assessment-for-Learning (Pre/During Learning) practices:

- (1) A teacher's experience teaching during Soviet-Era Uzbekistan versus post-USSR (such experiences seem to influence teachers' perceptions of the importance of Assessment-for-Learning measurements);

- (2) Being raised pre- or post-Independence (i.e., 1991) and attending an Uzbek, Soviet/Russian, or mixed-style secondary school – exposure to different types of classroom learning;
- (3) The number of years they have taught EFL (novice, mid-career, expert) and the number of professional development courses they have had;
- (4) Where they teach EFL in Uzbekistan (in the capital, Tashkent, or in the regions); and,
- (5) The amount of exposure they have had in Western/developed countries.

Throughout focus group and semi-structured interviews, I had pressed the participants to provide more detail about how they used assessment-*for*-learning techniques to inform specific aspects about the students' language learning and/or their teaching practices. In general, the participants were unable to explain, in English, perhaps because of their English proficiency levels, how they used the information they learned from assessment-*for*-learning techniques to support students' language learning. For instance, I asked several teachers to exemplify how they used class presentations to explain to students about specific language issues. The participants were unable to make clear if and how they used the presentations to support their students' English language development. This discrepancy between what they do and their reported cognitions shapes how they score students. A probable source of this discrepancy is that teachers may understand the purpose of an academic concept such as *feedforward*, but not know how to put the purpose into practice. In the second language teacher education literature, the sources of such discrepancies involve the differences between *academic concepts* and *everyday concepts*. Johnson and Golombek (2016) emphasize this difference using Kennedy's (1991) work:

...Kennedy (1991) who characterizes 'expertise' in teaching as emerging out of the ways in which teachers make sense of 'expert' knowledge, or knowledge that is propositional,

written down, codified in textbooks, and publically accepted as a principled way of understanding, phenomena within a particular discourse community (*academic concepts*), and their own ‘craft’ or ‘experiential’ knowledge that emerges through their own lived experiences as learners (*everyday concepts*) (p. 5).

As teachers begin to link the expert knowledge with their own base of understanding, they may be able to reformulate their everyday practices to include the academic concepts.

The participants reported similar procedures for how they kept track of students’ progress and eventually provided them with a score, regardless of their reported cognitions. Most participants assign students a score based on a general feeling. Only one teacher reported being trained in assessment and thus, regularly used rubrics for continuous assessment mechanisms. This teacher, however, was unable to explain clearly how she used assessment-*for*-learning techniques to inform students of their learning progress. Now, I turn to the second category that Uzbekistan EFL teachers use to assess their students: Assessment-*of*-Learning.

7.2 Assessment-*of*-Learning (Post Learning)

The second assessment category the participating Uzbekistan EFL teachers reported in order to assess students at the micro-institutional context is Assessment-*of*-Learning (Post Learning). This category “can be located at the summative end of the assessment continuum, where the outcome of assessment is to measure achievement on particular content” (Krause, Bochner, & Duschne, 2006, p. 857). The final examination is the category the teachers reported. The participating teachers’ comments on this topic may be further subdivided into: (1) how they design & evaluate final exams, and (2) how they administer and score these exams (see Table 31).

Table 31 Assessment-of-Learning

Broad Category	Sub Categories	Number of Teachers Reporting
Final Examination	Design & Evaluate	32 (60%)
	Administer & Score	15 (28%)

Note. 53 teachers in total.

7.2.1 Final Examinations

The participants reported they use a final examination to assess students. This test is administered at the end of a course of study and is meant to determine if a student has achieved the expected learning outcomes. The content of such final exams usually includes test sections dedicated to listening, reading, and writing, most often with a multiple-choice item format (see Appendix M for a sample final exam). The PRESETT curriculum, however, includes test specifications for each course, and often writing will be assessed with an essay. Additionally, the PRESETT curriculum tests speaking, in which a one-on-one interview format is used. Maimouna explained that few universities “[use] interviews for final tests” (Teacher 10_FG 2) because it is difficult to carry out in Uzbekistan. The amount of time to administer speaking tests is limited between the last day of class (i.e., at the end of the semester) and when grades are due. Trying to arrange teachers’ schedules can often constitute an administrative burden for teachers.

Maimouna reported that “[there are too many students] and [too] few teachers for a short time. A commission is needed to score speaking, but speaking tests are done.” (Teacher 10_FG2). The commission Maimouna refers to is constituted of a group of people, typically two teachers, who will score a student’s spoken language. Maimouna’s comment about “...speaking tests are done” (Teacher 10_FG2) emphasizes the point that, although there are often limited resources for carrying out speaking tests, they are in fact administered.

In Uzbekistan, a student can participate in a final exam only if s/he has successfully passed continuous and midterm assessments. Darisa from Focus Group 3 explained that “usually students [need to achieve a certain total score out of 100 points before they] can take the final exam. If they cannot gather that score, they will not be allowed to take the final exam” (Teacher 2_FG3). Teachers from all twelve regions confirmed this procedure. The total number of points a student can achieve on the final examination varies across schools and programs but is usually scored at 30 points. For example, at Feruza’s university in Namangan, the language department decided to separate the points differently among the final examinations, midterms, and continuous assessment, in contrast to most universities throughout Uzbekistan. The final examination for her university is the only one that was reported by the participants as having 50 points. Most participants reported that their universities have 30 points for the final examination. The weight of the final test thus varies across institutions and programs; however, teachers are not allowed to administer the final test to students unless each student has successfully completed his or her continuous and midterm assessments. Usually, between two to three weeks before the period of the final exam, teachers design the final assessments individually or in groups.

7.2.1.1 Test Design

Test design is the most common assessment practice reported by the participants. A test design practice common in Uzbekistan is that the department head decides if each teacher or a group of teachers within his/her English department will be charged with the design of the final exam papers. Once test papers are created, they are sent for initial review to the department head, a language teacher, or a testing specialist, and then distributed to the students. Uzbekistan EFL

teachers try to design their final examinations based on the curriculum requirements and/or the goals of the courses they teach.

7.2.1.2 Cognitions around Test Design

Some participants reported that the language test design process in Uzbekistan has many flaws. First, teachers do not feel comfortable designing tests because they often lack sufficient knowledge and skills of language assessment, and they feel that taking items (particularly multiple-choice ones) from the internet is an easier and more practical way to design tests. Second, the participants have different opinions regarding the potential effectiveness of the multiple-choice format. Multiple choice is widely used throughout many of the midterm and final examinations and has been identified by the participants as the most used test item format. Third, some participants question the validity and/or reliability of the kinds of final tests commonly approved by the administration. Fourth, some participants believe that the test process is unfair and does not truly measure what is being taught in all classrooms; the tests only reflect what is being taught in a few classes, and such a process is unfair to the students.

Some participants reported that they do not feel comfortable designing their own language tests because they do not have the necessary specialist knowledge and skills. Aisara and Ajva, who did not have a course in language assessment as part of their coursework for their Masters degree in TEFL, from Focus Group 1, discuss how they design tests.

Aisara: The head of our department [...] can say, “You four or five people...yes?”
 “[You all are] responsible for [making a] test.” [Then], [we each] have
 responsibilities.

Ajva: We can find it from internet or we can make it ourselves. Yes, very often from
 internets...

Interviewer: Can you please explain what you mean by ‘find it from internet?’

Aisara: So [...] we are not test developers. ... That's why we try to take some activities [of] reading and writing [...] from some books and internets. We are not going to develop it [ourselves] because we are not experts. We just take [from the internet]. Of course, sometimes, [the contents] [are] connected with the material which we just [taught]. They are not the same [material, but, ones that we adapt].

This discussion among Aisara, Ajava, and me illustrates that the department head has considerable control over who designs tests. The participants reported going onto the internet to find items, reading passages, and/or questions that would best connect with what was taught in the course materials, because they did not feel comfortable writing their own assessment items. Aisara in Focus Group 1 stressed that she is not a specialist test developer and felt uncomfortable creating test items from scratch. At the end of the focus group interview she said that she would “hope to learn more about the science of tests” (Aisara_FG1). Many teachers revealed they are responsible for designing tests but feel that they have insufficient background knowledge and/or skills to serve as test developers.

Most of the participants commented on the practice of using multiple-choice items on their final examinations because multiple-choice items seem relatively “easy to do and really easy to score” (Abdulaziz_FG2). However, Diora from Khorezm, who is a novice EFL teacher, has been reading extensively on recently published TEFL methodology and testing. She “tries to use [the multiple-choice format] as less as possible on the tests. [Because] ... [students can find] the answers from some books [and] internet and like this. [So,] I try [other formats other than multiple-choice] ... For example, I give [students a] passage from one novel, and they should

guess [the writer's] attitude, [and write the answer in prose form]...Yes, of course, I use multiple-choice tests, but not very often" (Diora_4). Diora believes that having students produce language (e.g., writing) would reveal to her more reliable information about their language abilities than the items of a multiple-choice test.

Unlike Diora, Almina is an older EFL teacher in Uzbekistan and has been teaching English since before the independence of 1991. Her past experiences in language test design have influenced the way she views the multiple-choice format. While participating in Focus Group 2, Almina reported that the students need to be able to do *any kind of test method* because the tests are grounded in the knowledge of the course. She explained that, "I know. I gave the information. I gave the knowledge, and according this information, the knowledge, I prepare my tests. I don't care [whether or not they can] manage this, could they solve this test or no. I know that I gave them this information, and they should know this..." (FG2). Almina is the only person who reported that she *does not care* if a test item is too difficult (or easy), because she believed the students should have learned all the information she provided for them during the course. Thus, she does think about the test method effect and how that could have a role in a students' performance.

After the teachers create their tests individually or in a group, they then send the test papers back to one of three parties: the department head, a particular teacher selected by the department head, or a language testing specialist for checking. Teachers who reported being a part of the PRESETT or ESP curriculums have similar experiences in the test development process and believe that after all their test papers have been collected, then, a testing specialist will read them over and edit.

I will compare the experience of Mukaddas (from Bukhara), who teaches within the PRESETT curriculum, with Nozliya, who is an ESP teacher from Syrdara. They both have had similar experiences with the process of test construction. Mukaddas explains her test development procedure:

So, in our department, what we do is that each teacher, before the assessment starts, we have the time period we have set. We design tests, two or three, usually three test papers. [We] collect all of them in our department and one very skillful teacher is appointed by the head of the department to edit and to check the quality of these tests. So she or he makes sure that all the papers, tests, are good to go, good to use in the examination. Then we use these tests. And again, we need a number of tests and variety and we want to make sure we have a lot of different papers (Mukaddas_12).

This test development process gives some degree of autonomy to each language teacher so that he or she is able to participate in the creation of an examination that accurately reflects what has been taught in his or her classes. Nozliya is an ESP teacher from Syrdara, and she has a similar experience to that of Mukaddas:

Our head department gives us an assignment. She asks us to make tests. Let's say this is for midterm, so midterm, final, same way. Before the exam, maybe two weeks before the exam, all of the teachers are given assignments to bring tests and she gives us a sample test, so the test should be this, multiple-choice or whatever, and she shows us the criteria and the level that we need to consider, and we take the sample test and we try to make the tests ourselves. So each teacher brings five different [versions of] tests, and we all collect those tests and usually there's one teacher who is responsible for editing and checking for mistakes. [Then], before the final exam, all the tests are gone through, edited, and

corrected [just in case] there are [any] mistakes, and then this is how we distribute tests (Nozliya_11).

Both Mukaddas and Nozilya explain that a testing specialist, or a qualified language teacher, is appointed by the department head to review and edit the exams. However, Nodira from Andijan – the easternmost region of Uzbekistan – rejects the idea that a person checks the quality of the tests at her university, and does not believe that this process happens across Uzbekistan. She explains that she “doubts [they follow such procedures because the tests that I see and I am now involved in are very poorly written. I am afraid that we are not measuring accurately. Maybe we are measuring 50% accurately. We have practicality and validity issues of the tests we give our students. I doubt that they were reviewed by someone. There are so many mistakes in the questions and sometimes there are no answers They are not of good quality, I guess. It's my idea, but not only mine. But many teachers think so]” (Paraphrased for clarity, Nodira_6). Nodira has taken a U.S.-based language testing course and uses her knowledge of testing principles to question the final tests that are distributed to students.

Other participants commented on evaluating the final examination. Below is a conversation I had with Ulugbek from Djizzak:

Ulugbek: After tests are given [to students to take], I have seen many tests that [I believe] are poorly written, many times, and I say something.

Interviewer: Tell me more about that?

Ulugbek: I have seen mistakes in the test [papers], which were not edited properly before the exam. I have seen [problems] in those [test papers which were created by other teachers]. [If a problem in the test paper was recognized] we would not lower the score of a student, because, it was our mistake, [the] teacher's

mistake, and we [will] give [the student] one extra point [...], so we add scores for this particular question which contains a mistake, so I have seen these kinds of problems.

Interviewer: When you bring up these problems, do you as a group, I guess with the people that make the exam, talk about how to fix it for next time?

Ulugbek: Mm-hmm (affirmative). Yes, we do. I think we should do some math [conduct some statistical analyses of some test items], but no one really knows [no one is available who knows how to do this.]

Ulugbek points out that the final examinations created for his university have problems. He believes the exams have poorly-written items because they were not edited properly. He wants to learn how to conduct item analysis, but he lacks the training and resources (e.g., materials) to acquire such knowledge. Thus, Ulugbek is following his belief that the tests are poorly constructed, and he feels that by adding more points to students' scores he is making up for how poorly constructed certain teacher-crafted assessments are. My conversation with Nodira from Andijan matches Ulugbek's intuition about when she reviews the final examinations for her department.

Interviewer: How do you know if your tests are actually measuring what you're trying to measure? Let's say I'm teaching writing. In my classes, I'm doing X, Y, Z. How do you know your exam actually tests X, Y, Z?

Nodira: We only assume. We hope.

These sentiments were echoed in all of the focus groups. Most Uzbek language teachers only hope the tests are valid (i.e., measuring what they are intended to measure).

Most Uzbek language teachers report they had problems with how schools ask teachers to make tests. Many participants said that they want to test their students the way that they want to test them, and not feel that they should have to have their tests be approved by others. Mukaddas explained the controversy:

Sometimes the challenge is that some [test] papers [are different from each other]; for example, I have designed mine, you have designed yours. [Each teacher also has] a different style in teaching and [the make-up of our classes are also very different]: The [students'] levels are different, the tasks [that we use for the tests] are different. So, [after students take the] test, [some] come and ask [me], 'Okay, how come my friend had this task, and I didn't?' So that is the challenge. If I use pre-intermediate level texts or tasks [for] my test sheet that I have designed, [and then] my colleague might think about her lower level students and [use] lower level materials [for her tests] [then, we are not giving equal test papers to students.] So, [when] we mix these test papers [as a group], her papers might be used for my [classes] that I teach. My papers might go to her classes who are lower level students. [Do] you understand what I'm talking about?

Thus, the number of test papers increases in a larger department because they will have more students and a wider range of proficiency levels-of-students. According to Dana in Focus Group 5, "the chance of cheating lowers [i.e., lessens] when more test papers are created" (Teacher 5_FG5).

The Uzbekistan EFL teachers have varying cognitions about the test design process. Some believe that they themselves are not strong enough in language assessment test design and construction to be able to create effective tests. Others believe that the type of test items do not matter because the students should be able to perform well on any items that measure what

teachers perceive they should be tested on. A few participants reported that the process of a testing specialist who reviews tests is questionable, because they have identified many flaws in the test papers, and cannot believe that a specialist could oversee egregious errors. Overall, the participating Uzbekistan EFL teachers would like more autonomy over the test design process so they are able to assess their students the way that they want to test them. Additionally, some participants realize that they lack certain knowledge and skills of language assessment construction and evaluation and would like more training on how to properly design, administer, and evaluate effective language tests.

7.2.1.3 Administration and Scoring

Fifteen participants reported on how they administer and score final examinations. Gabriella in Focus Group 2 has had positive experiences and explained that the process of test administration is "...very interesting, at school, [the] final [test] must be taken by another teacher [and that teacher also] gathers the [scores] ... for example, I am a teacher of listening and speaking, yeah? My final must be taken by another teacher and checked by another teacher. I [am] only [allowed to] see the results and write [them] down on the paper [to submit to the school]" (Teacher 5_FG2). Gabriella explained that in Uzbekistan each teacher is not expected to give the final examination paper to his or her own class of students. They are expected to distribute the tests to another class, and then score that class's papers. Farrukh (Teacher 9) and Sakina (Teacher 11) from Focus Group 4, as well as all one-on-one interview participants, confirmed this administration procedure.

All Uzbek language teachers are instructed to score the final exams of a colleague's class and then provide the results to the teachers. No teachers reported they score their own class exams. Maimouna in Focus Group 2 explained that he sees "Only the results, I see only results. I

couldn't check-up, I couldn't take an exam from my class, and I could see only results and say your results they are real it's obvious, bad or..." (Teacher 10_FG2). The procedure of language teachers having to scoring another class's test is echoed by teachers from each region throughout Uzbekistan.

7.2.1.4 Cognitions around Test Administration and Scoring

The participants believe that the test administration and scoring processes are difficult for teachers who are not content-area specialists in the courses they are scoring. Klara from Tashkent discussed the difficulty of checking someone else's class examinations when one is not the content-area teacher for that section. Klara teaches speaking and listening to English Language and Literature majors who are part of the PRESETT curriculum; however, she was directed by the department head to distribute tests designed to assess students' abilities as writers.

Klara: Writing is more difficult. In one group, maybe twenty-five students, we have twenty-five exercise books. That we're checking all of them...

Interviewer: When you say, twenty-five, twenty-five ... You mean twenty-five students?

Klara: It's a lot of students. I have only three days to [check] these ... Yeah, it's only three days. You feel rushed to check all of that and I don't teach writing.

Interviewer: Mm-hmm (affirmative).

Klara: I don't want [to feel rushed] ... Anyway, it's difficult [for] me. That's why [it will take me the] whole day [and night to score]. I have only three days to [score] these notebooks.

In most universities, it was reported that teachers who score speaking and writing used rubrics to provide grades for students. However, the teachers who do not create these rubrics often have a difficult time using them to provide students with accurate scores.

7.2.1.5 How Students are Scored on Final Examinations

From the onset of the final examination test design process to the administration and scoring of exams, the Uzbekistan EFL teachers reported they become more removed from their own students and possibly the content area they are accustomed to teaching and working in. Many teachers have identified negative feelings about the final examination and do not think the tests are effective in assessing the required knowledge and skills of the courses they teach. Teachers who talked favorably about the final examination test design process, however, were teachers who had a direct impact on what tests (i.e., formats and items) were used with the students that they knew could do well. How much direct responsibility a teacher had in the process of test design, administration, and scoring contributed to the teachers' cognitions and emotions. Most times though, the teachers were asked to score students who they did not know, and in a content area that they did not work in, with a scoring mechanism they were not accustomed to or were untrained in. The lack of connection to students and responsibility, with negative emotions, contributed to scoring that led to evaluating from ill-defined impressions rather than from the use of specific criteria.

For instance, many of the teachers tend to disregard rubrics and score students based on their overall impressions of the students' writing or speaking abilities, possibly because they were not trained in the use of rubrics for assessment purposes. Nozliya from Focus Group 5 explained that she tried to "look at the criteria," when she scored writing. But, she said, "sometimes I can't [use the rubric]. Maybe I like this essay, for [it's] style, but there are many

grammar mistakes. [Then,] I change my [personal scoring] criteria because I like this one's style [and judge other students' writings to it] I support [this] student's [score as high]. I can close my eyes on the grammar mistakes" (Teacher 2_FG5). Nozlyia from Focus Group 5 emphasized that her impressions play more of a role in the scoring of writing than on what is considered important on the rubric.

Evaluating writing and speaking according to loosely-defined impressions is reported often by the participants. Umida from Karakalpakistan, who oversaw the test design process at her university, recognized this tendency of Uzbek teachers when they score writing, and explained that at her university, "people are more grade and rubric-oriented people. [They do not take their own more impressionistic appraisals of the quality of the student's work into consideration.] Our teachers [score differently from many teachers in Uzbekistan because] we have [a] better understanding [of] language learning" (Umida_5). Thus, Uzbek EFL teachers vary in how they score writing and speaking – some teachers adhere to rubrics, while others tend not to do so. The difference among the Uzbekistan EFL teachers participating in the present study was that those who had a more direct responsibility in the design, administration, and scoring process of the tests tended to perceive existing test design processes more favorably, while those who were not part of the decision making viewed the processes less favorably.

7.2.2 Conclusion – Assessment-of-Learning (Post Learning)

The Uzbekistan EFL teachers reported Assessment-of-Learning (Post Learning) techniques as one of the assessment categories they employ at their institutions to assess students. Assessment-of-Learning (Post Learning) is meant to measure what was had been taught by the end of the semester. The Uzbekistan EFL teachers reported they extensively used a final examination for testing listening, reading, and writing. These tests often use the multiple-choice format, while

some writing exams also incorporate essays, and speaking tests often use the format of a one-on-one interview.

The participants have varying cognitions about Assessment-*of*-Learning (Post Learning) techniques, particularly the final examination development process. First, the data revealed that most of the participating teachers do not feel comfortable designing tests because they often lack requisite knowledge and skills of language assessment. These teachers questioned their test development abilities and felt that using resources from the internet was an easier and more practical way to design tests than develop tests from scratch. Second, the participants have different opinions on the effectiveness of different test item formats (e.g., multiple choice); some do not think that the format accurately measures language, while other teachers believe that if the student has learned knowledge, then the test-item format does not matter. Third, some participants do not believe that the final tests approved by the administration represent adequate measures of students' language knowledge and attainment. Fourth, some participants believe that the test development process is unfair and results in testing instruments that do not accurately measure what is being taught in their classrooms; the tests only reflect what is being taught in a few classes, and, therefore, are unfair to many students.

There are several different factors identified that may influence Uzbekistan EFL teachers' cognitions about assessment-*of*-learning practices:

- (1) Educational experiences in an MA-level TEFL program, and whether the program featured a language assessment/testing course;
- (2) Amount of time training (e.g., professional development courses) in language assessment/testing;

- (3) Experience using certain item formats to test students' knowledge about language and language use;
- (4) Experience using certain scoring procedures; and
- (5) Reading up-to-date literature and research on language teaching and testing practices.

Most times, though, the teachers were asked to create exams that were not used by the groups of students they taught. Also, they reported that these exams were either not edited at all or they were poorly edited. The participants were told to administer and score exams for students who they did not know, in a content area that they did not work in, with a scoring mechanism they were not accustomed to or trained in. The lack of connection to students, which produced negative emotions, contributed to the way the Uzbekistan EFL teachers scored students. As a result, the scoring procedures they followed tended to be highly impressionistic/subjective.

7.3 How do Uzbekistan EFL Teachers Justify Assessment Practices?

This chapter addressed the following research question: How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students? To answer this question, I asked the following three subquestions: A. What do Uzbekistan EFL university teachers report they do to assess their students (i.e., assessment tasks and scoring procedures)? B. What are their cognitions surrounding those assessment practices? C. How do their cognitions about assessment shape how they assign scores to their students' work? As assessment is inherently an inferential activity (Fulcher, 2015; Lado, 1961) and involves the collection of information to make a decision that leads to future consequences for a person or a group of people (Inbar-Lourie, 2008a), the results featured in this section question whether Uzbekistan EFL teachers create and use assessment experiences for the betterment of students' learning.

As noted throughout the chapter, the Uzbekistan teachers conducted Assessment-*for*-Learning (Pre/During Learning) – activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged (Black & William, 1998). The participants reported conducting continuous assessment and employing placement/diagnostic assessments practices. Additionally, the participants utilized Assessment-*of*-Learning (Post Learning), which in the ideal world would be based upon sound psychometric testing conceptualizations and practices, and measures how much knowledge has been learned over time. The purpose of summative assessments is aimed at examining the extent to which knowledge (e.g., the material covered from the curriculum) has been acquired. The teachers reported participating in the design, evaluation, administration, and scoring of final examinations.

Although the participants reported *doing* Assessment-*for*-Learning (Pre/During Learning) and Assessment-*of*-Learning (Post Learning), they generally did not justify how they score students to support students' learning, although such was the intention of many of the participants. There was a discrepancy between what the participants reported they did, what they thought about what they did, and how they explained the scores they assigned to students. For instance, the participants reported that they think, believe, and feel that certain assessment procedures (e.g., continuous assessment) are helpful for identifying and helping students with the challenges they face as language learners. However, these participants were unable to explain how they used the information they gleaned from the continuous assessment (and all assessments) to inform students about ways they could improve. In addition, many of the teachers reported that they scored the continuous assessment on general impressions (i.e., rather

subjectively). Also, teachers who did not think that continuous assessment was helpful still graded their students because they had to provide them a score. These teachers who reported more of a negative feeling about updated assessment practices were deeply impacted by the education system of the USSR. The USSR system often resisted different ideas about assessment. These teachers also scored their students subjectively. In general, both types of teachers – those who had positive and negative feelings toward continuous assessment – scored the students on the language teachers' general impressions. One of the participants, Yayra, explained that this type of assessment practice is slowly being changed in Uzbekistan. Her remarks come from a space in which she has been steeped in both assessment cultures – she was trained in language assessment in a Western country, and she has been teaching EFL in Uzbekistan for fifteen years.

The participants generally noted that they are unsure if the tests and all assessment practices were measuring what they were trying to measure. I will reiterate Nodira's comments about the meaningfulness of the participating Uzbekistan EFL teachers' assessment practices because I find it a good way to summarize if the assessment mechanisms are meaningful. When I asked Nodira about how she knew if her final examinations were valid, she replied, "We only assume. We hope" (Teacher 8_FG3).

7.4 Discussion

I begin this discussion by referring to the description of the participants in Chapter 4 and present the descriptive statistics, which will put into perspective the participating Uzbekistan EFL teachers' cognitions around language assessment. (See Table 32.)

Table 32 Descriptive Statistics of Teachers' Age and Years' Experience (in years)

	Mean	Median	Mode	Min	Max	Range
Age	37	37	39	26	63	37
Years' Experience	11	14	15	1	40	39

The descriptive statistics, on average, show that the participants' ages cross the year of 1991, when Uzbekistan gained its independence from the USSR. Fifteen of the fifty-three teachers reported teaching English and/or Russian for over 25 years and thus taught language during Soviet Era Uzbekistan. The participants were language teachers during this historical transition while others were language students in middle schools, high schools, and universities across Uzbekistan. Lortie (1975) noted that teachers' schooling and professional coursework contribute to teachers' background and the formation of their thinking. Thus, for some participants, their cognitions around language assessment and testing are grounded in what was understood as effective and meaningful during the period of USSR influence in this part of the world.

These cognitions were reported by the older generation of teachers in the focus groups and semi-structured interviews. Sometimes, they “did not care” (Almina_FG2) about varying testing practices, because they believed that if students learned specific knowledge and skills in their classes, then they should succeed on the tests being administered. This educational belief is commonly held by Soviet-Era educators (personal communication, Azizov, May 25, 2018). These teachers were less open to more contemporary ideas about assessment that originate within Western countries (e.g., continuous assessments). Some did not understand that participation could be broken down into different criteria because the older generation of teachers generally reported that it is the students' job to come to class, because there is no other way for a student to learn. These teachers identified their role as an information-transmission conduit of information about language that they passed down to students. Additionally, it should be noted and/or recognized that most teachers who taught languages during the time of the USSR

influence on the educational system of Uzbekistan taught Russian rather than English and then were forced to teach English at their universities. This change of language could have contributed to a negative impression of the English language and/or the cultures of English-speaking countries. Generally, the older EFL teachers held more entrenched beliefs about language assessment that was viewed during the time of USSR's control of the education system in Uzbekistan.

Younger and less experienced language teachers were introduced to the contemporary role of the English language in business, education, communication, diplomacy, tourism, and entertainment during the era of Islam Karimov (see Chapter 3) (i.e., 1991-2016). These teachers participated more frequently in updated practices of language teaching and learning, and all received their MA degrees in English language teaching from universities across Uzbekistan. These second language teacher education programs often focused on the communicative approach and stressed social dimensions of language learning. Thus, the language teachers who participated in professional coursework designed from this perspective of language learning and teaching reported favoring the use of different assessments that focus on interaction (e.g., presentations). If a novice teacher was educated prior to 2015, then he or she would not have been exposed to the PRESETT curriculum, in which language assessment was first introduced as a key area for pre-service language teacher education. If a participant was trained in the PRESETT curriculum, they are more likely to have been exposed to contemporary tenets of language assessment and testing and are more likely to have a positive impression of the value of *Assessment-for-Learning* and *Assessment-of-Learning* concepts and instructional practices.

All participating Uzbekistan EFL teachers have cognitions about language assessment, which vary from thinking that certain assessment practices are valuable and worthwhile to

improve students' language learning to not needed and therefore not necessary. There are five overarching factors identified in this chapter that could influence the participants' cognitions about language assessment practices:

- (1) Teachers' experience teaching (or learning languages) during Soviet Era Uzbekistan versus post-USSR;
- (2) The number of years in their career (novice, mid-career, expert) and where they have taught EFL (countryside or city), and the number of professional development courses in which they have participated;
- (3) Educational experience in an MA-level TEFL program, and whether it had a language assessment/testing course; amount of time training (e.g., professional development courses) in language assessment/testing;
- (4) Experience using certain item formats to test students; experience using certain scoring procedures; and
- (5) The amount of exposure they have had in Western/developed countries.

As identified above, these five factors (with some overlap – e.g., 3 and 5) correspond to Borg's (2003a) conceptualization of language teacher cognition (see Chapter 2). The Uzbekistan EFL teachers' schooling, professional coursework, contextual factors, and classroom practices (Borg, 2003a) have an impact on their cognitions surrounding language assessment. This finding is similar to that of other research that investigates different aspects of language teacher cognitions (see Chapter 2).

From classroom teachers' perspective, to justify using assessment practices at the service of learning, one would need access to procedural knowledge and know how to apply knowledge about language assessment (both declarative knowledge and procedural knowledge) in relevant

ways as part of language instruction. The study's results signal that the participating teachers, although they have declarative knowledge about assessment principles [e.g., positive washback – the positive influence a language test could have on the teaching and learning of language (Cheng, 2005)], are unable to put that knowledge into practice and tend not to be able to utilize assessments for learning. Thus, there seems to be a disconnect between what the participating teachers do, what they think about what they do, and how they assess students for learning.

A major limitation of these findings is they are based on the participating Uzbekistan EFL teachers reported assessment practices about, thoughts surrounding, and responses to how they score their students with *Assessment-for-Learning* (Pre/During Learning) and *Assessment-of-Learning* (Post Learning) measures. To confirm if what they reported and discussed is indeed, true, I will eventually need to move beyond data collection procedures of focus group and semi-structured interviews to classroom observations, observing teachers engaged in the process of assigning scores, and conducting document analysis of midterm and final examinations. There are many more research methods available that one can use to corroborate the current findings. The close examination of practicing teachers' reported assessment practices and cognitions is a start to discover the language assessment literacy of Uzbekistan EFL teachers, and can inform language teacher educators who specialize in language assessment, the means to assist Uzbekistan EFL teachers to become language assessment literate. (See Chapter 9 on implications for teacher education.)

8 RESEARCH QUESTION THREE RESULTS AND DISCUSSION

I present the findings of the qualitative analyses, which were conducted for the following research question: What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students? To address this inquiry, I asked the following three questions: A. What are the macro-environmental factors Uzbekistan EFL university teachers report? B. What are their cognitions surrounding these factors? C. How do the reported factors and cognitions shape Uzbekistan EFL university teachers' assessment practices?

To answer these questions, I used the same methodology from Chapter Six, in which I interviewed forty-one EFL teachers who participated in focus-group and/or one-on-one semi-structured interviews. I then conducted an analysis of the transcribed interviews to answer questions one, two, and three. Throughout five focus groups and twelve one-on-one interviews, the participants reported two external factors, which were broadly categorized as sociocultural and sociopolitical. To check if each coded category could stand on its own, I used 15% of the data and conducted the card-sorting technique (as cited in, Nunan & Bailey, 2009 with one other rater. The rater reliability for the following categories was $\alpha = .89$. I will begin with a discussion of the sociocultural factors and then the sociopolitical factors identified by the participants. I will discuss each aspect with relevant quotations from the focus-group and semi-structured interview data.

8.1 Sociocultural

The first reported external factor is *sociocultural*, which seeks to understand why people act as they do based on the influences of their social and cultural group memberships. The participants

discussed three specific topics within a sociocultural perspective that shape their assessment practices. The first topic is the cultural value of collectivism, the second is the perceived role of a language teacher in Uzbek society, and the third is Uzbekistan EFL teachers' ethical responsibility. I will discuss each of these in turn.

8.1.1 Collectivism – “Strength is in Unity” (Karimov, 1991)

The participating EFL teachers identified their culture as a collectivist society, which means they value group relationships over individual pursuits. The idea of the self is construed as an interdependent entity, similar to other Central and East Asian cultures (e.g., Japan and China). Madison from Focus Group 3 explained that the “Uzbeks have such kind of value, it comes out of our mentality, cooperation ... If we unite together, we will develop” (Teacher 4_FG3). Parizoda in Focus Group 5 revealed why she believed that Uzbek culture is considered collectivist: “In Uzbek culture [we] do everything together, because [being together] makes everything easier, cheaper, and better” (Teacher 8_FG5). The collective nature of Uzbek culture is also rooted in the language. Klara explained that “Uzbek has several words to show united activity, such as *hashar*. [This word is used when we all come together to do something for free, such as clean streets, build houses or roads with each other. *Tuyona* is another word that we use, which is the money we give for a bride and groom at their wedding]” (Paraphrased for clarity, Teacher 2_FG1).

A hallmark of Uzbek education is the common belief that everyone, regardless of one's societal or economic status, should come together for the greater good. Additionally, the participants characterized university departments as a second home, because teachers spend most, if not all, of their day at school. In the first focus group Klara said, “You know, the department is one family because we work 80% of the day [together]. I [see my children less

time than my colleagues]. [So,] we must be a unified group. Our purpose is to help students” (Teacher 2_FG1). Klara’s comment underscores the importance of the relationship a language teacher could have with his/her department and students, and sees this interconnectedness as family-oriented.

8.1.1.1 Cognitions around Collectivism and Assessment

The most frequently reported cognition around collectivism and assessment is the commonly held belief that *students should not fail*. This practice means that if a student is unsuccessful and fails to pass a course, then s/he will have an opportunity to complete the class at a later point in time (e.g., winter or summer holidays, later years). Zumara and Samia from Focus Group 4, who teach ESP at the same university in the autonomous region of Karakalpakistan, explained the general situation:

Zumara: In general, all of them pass. All of them pass...

Samia: ... ya... We have such kind of law in our university. Law, which was made from our faculty ... And they are written that students have [as many chances as they need to pass the course].

Zumara and Samia showcased how their institution explicitly created a written rule that they are unable to fail any students. The rule was created by Karakalpak and Uzbek faculty in the region of Karakalpakistan, which is the northwestern most province in Uzbekistan (see Chapter 3 for region map) and is identified by the Uzbek government as an autonomous region. Zumara emphasized that “[The rule was created because] students at our university were born in Nukus or surrounding area like Akmangi and their families are poor and only make a small amount a month” (Teacher 1_Focus Group 4). Zumara’s explanation of her students directs our attention to

a description of the type of student body at her university. Most students are Karakalpak ethnic born while some of them are Uzbek and Kazakh. Most students come from poverty-stricken families. Zumara feels that the decision not to fail students is good for the society in that region of Uzbekistan, because they have been through tough economic times, and she feels that she should be supporting students instead of failing them (Teacher 1_Focus Group 4). The written rule of not being able to fail students was supported by many, but not all, participating Uzbekistan EFL teachers.

In contrast, Nozliya from Syrdara who teaches in the PRESETT curriculum and who is Russian-born and educated, views the axiom that “students should not fail” as antiquated. During our one-on-one interview, we had a conversation about her experience failing students and the cultural backlash she received from her Uzbek colleagues:

Interviewer: Have you ever failed a student?

Nozliya: Yes, I have. Last year I did.

Interviewer: Tell me about that experience.

Nozliya: A couple of them, like four, I had four students who failed, but I didn't give them a chance, and of course, head of the department first asks the teacher, "Should we give them a chance? What do you think? What do you see the student, if we give them a chance on it?" I said, "No, there's no point because all year, for one year, the student didn't do good, and how is he going to ... " I didn't see a student progressing when I gave him a chance, so I did not, and the head of the department also didn't approve, and so the student failed, and they came back in September to retake the course.

Interviewer: How is that seen in Uzbekistan? I'm sure everyone knows if you retake a course.

Nozliya: It's not seen as a good thing. It's not a good thing, for both student and the teachers. Teachers are not liked or welcomed when they fail students. I was not liked.

Interviewer: Not liked by who?

Nozliya: By other teachers, other colleagues around us.

Interviewer: Can you explain more about your experience?

Nozliya: Mm-hmm (affirmative). They will try to influence me, so, "What are you going to gain out of this?" So students are coming from regions, they're in difficult situations away from home. "Just help them out."

Nozliya and the head of her department believed that the students who failed Nozliya's course failed because they did not do the required work and/or achieved the appropriate level/requirements of the course. Although she went through the proper channels to not pass her students, she did receive negative backlash from her Uzbek colleagues. They viewed Nozliya as not being part of their collective community, nor as someone who holds the same beliefs that they do, and therefore she was not very well liked. The Uzbek teachers wanted her to "just help them out" (Nozliya _11) and Nozliya identified this aspect as societal pressure. Nozliya, however, would not change her opinions based on the pressures from her colleagues. Thus the Russian-born English teacher Nozliya, and the Uzbek EFL teachers view assessing learners differently, while the Uzbekistan-born EFL teachers placed more emphasis on supporting the community over the individual.

8.1.1.2 Collectivism Shaping Uzbekistan EFL Assessment Responsibilities

The participating Uzbekistan EFL teachers assess their students through the perceptual prism of collectivism, which plays a role in the different assessment decisions that they make, such as influencing how they choose and develop assessment procedures. Additionally, the participants identified their role not only as a language teacher, but also as a parent (Umida_5). Darisa from focus group 2 emphasized that “The [Uzbek language teachers] want to have an emotional approach, cultural approach toward assessing... Meaning they want to feel ... What's the word? They are sorry for students if they do bad and happy if they do good” (Teacher 2_FG3). In general, the participants want the best for their students, and they will try all that they can in order not to fail them. Teachers' assessment practices were reported to be affected by the cultural value of collectivism, especially when students' personal circumstances were impacted by pregnancy and/or marriage. Such personal circumstances are important in Uzbek society because they relate to the concept of family.

8.1.1.2.1 Personal Circumstances – Pregnancy and Marriage

A student's pregnancy could determine if a teacher 1) provides more time for the student to complete the coursework and exams; 2) rearrange the requirements for how the student will be assessed; or 3) only produces a final score for the student in the class (thus, producing an automatic pass). In Focus Group 2, Diora explained why a pregnant student needs more time to complete the course and/or examination:

Diora: Yes, she's pregnant. Yes. I don't think that there are some kind of- we should not be so strict. We just try to be a bit [more] flexible.

Interviewer: Flexible, yeah.

Diora: I usually, for this kind of student, I will say, "Okay, I know there is a fixed date by the Dean, but you may come [on some other day that we can arrange]. Whenever is good for you."

Diora provides more convenient time for the student who is pregnant because she recognizes how difficult it is to carry a child. Her opinion echoed that of Ona from Focus Group 3, who explained that all of the teachers “are trying to be honest teachers, [without bias], who will try to assess the learners, the knowledge of the learners not their money, not their relationship, not. Yes, but sometimes we of course consider if there is a pregnant woman who has this kind of- It's very hard to carry the baby and she needs energy” (Teacher 6_FG 3). The importance of pregnancy, and the difficulty of some women to bear a child, is taken into consideration by the participating Uzbekistan EFL teachers.

One participant, Nodira (with whom I discussed the subject of pregnancy during the semi-structured one-on-one interviews) experienced pregnancy herself during her university career and she wanted to talk about what it was like to carry a child while attending school. I asked if she could explain what it is like to be a teacher now and have students who are pregnant. In response, she related the following anecdote:

Once my student was pregnant, and she couldn't attend several classes, and she gave birth to her child before the [due-]date, and her health suffered because of it, her child and

herself. Even then, she came and discussed the situation with me. [She even came directly] from the hospital [to talk to me] and so, then I changed it [to pass]. (Nodira_6)

Nodira provided her the passing grade because she trusted the student who said that she was sick at the hospital with the child. Nodira made a choice to help her because she could relate to how difficult life can be when bearing a child. The birth of a child is highly valued in Uzbek society, and thus reveals a large contextual feature that can influence if and how a teacher follows more standard assessment practices.

In addition to pregnancy, marriage can affect the amount of time students have available to complete assignments and tasks. In a one-on-one interview, Aziza described the importance of marriage to a typical Uzbek family, “As we know, as soon as a child is born, parents start preparation for his/her marriage by collecting *sarpa* [all necessary household items] for a girl or starting to build a house for a boy so that in twenty years everything is ready for a child to get married” (Aziza_7). According to Aziza, a cultural goal of an Uzbek is to continue the traditions and practices of Uzbek society, which first and foremost begins with marriage (and subsequently having children). Thus, students who get married while a course is in progress would receive more time to complete classroom assignments and tasks including class projects, papers, and/or homework assignments. Usually, according to Parizoda from Focus Group 5, “Marriage, in marriage we can give [overlapping speech]. Yes. Two, three days” (Teacher 8_FG5). All members from the fifth focus group confirmed Parizoda’s comment, but in the teachers’ overlapping speech, they were discussing the number of days/months that students who get married would be able to complete class projects and papers. Some teachers uttered 5 days, 1 month, and even one year. In essence, the amount of time varies across universities, but the central idea that more time is provided for students to complete assignments.

To summarize the value placed on pregnancy and marriage, Aziza explained that “During the student years it is okay (and even welcomed) if students-girls get married and have children. All teachers, and administration support it and approve it, even if the girl missed many classes ... the most important is that she *got married* (emphasis added)!” (Aziza_7).

8.1.1.2.2 Failing Scores and Other Avenues

Most participants, who want to be accepted socially by colleagues, parents, and/or students, reported that they know they are unable to fail students. Thus, if a student is not successful at passing the final examination, many participating teachers reported that they also consider the students’ work ethic and natural language skills before providing a failing total score.

8.1.1.2.3 Examining Students’ Work Ethic and Language Abilities

One of the ways the participants report that they assign final scores at the end of the semester is based on the student’s work ethic throughout the entire semester. Teachers might ask themselves, ‘Did the student complete all assigned readings/tasks, participate in class discussion, and come to class on time’? Diora, from interview 4, commented on one of her students who had done exceptionally impressive work throughout the whole semester, and that therefore it was difficult to give her a lower score when she performed lower than expected on a test. Diora reported that “[for overall assessment, she did well]. Just the final exam, she had a problem. And I took that into consideration. Mm-hmm (affirmative). [I] Gave her [a] high score then, not too high, but good” (Diora_4). It is not clear whether Diora took all of her student’s work ethic or their previous performances into consideration when she assigned final scores. Some of her students made stronger impressions on Diora than others.

Bibidana in Focus Group 3, supports the idea that some teachers are more aware of some students’ successes and failures throughout the entire course of study. Bibidana explained that

“For each student, I have [my] own [way to track their work style throughout the course]. [I do not] get [ideas] from other teachers. [If a student makes some mistakes on a test, then,] they made some mistakes. [But], [I pay more attention] to his study during the whole semester” (Teacher 1_FG3).

The second way some of the participants report that they score their students is based on students’ language abilities, regardless of what they have learned in the course. That is, a student might have strong spoken English skills because she had lived abroad in an English-speaking country for many years. This student’s spoken ability could have an influence on some teachers’ assessment practices. For instance, Aziza explained that “Mostly they [Uzbek teachers] refer to speaking abilities. It doesn't matter if even they're teaching writing. For example, if I am teaching writing, there are some students who have been in [the] U.K., in [the] U.S., and they come, but they don't have appropriate writing skills... [So, teachers will pay attention to his spoken ability when they score his writing. They might ask themselves, ‘How can I give him a low score when his speaking is good]? [Teachers normally] do not refer to the piece of writing [when they score]. [Well], most of them” (Aziza_7). Many of the participating teachers seem to hold the belief that spoken language proficiency is the ultimate level of language achievement, and, if students have advanced spoken language abilities, then, they are likely to achieve higher scores during their coursework.

If an Uzbekistan EFL teacher does not pass a student, which is rare, the general procedure available to all students at all universities is as follows: If a student fails a class, he or she must obtain permission from someone such as the department head, dean, or vice rector of the university to retake the final examination. Once permission is secured, the student can retake the same test (e.g., same content and test items). At this point, accepted procedures are that the

classroom teacher is not involved in the process of designing, choosing, or selecting materials to assess the student. Having alternative formats for final examinations is meant to eliminate cheating during the examination period; however, toward the end of the course, if a student is unsuccessful, then, the same format is meant to make it easier for the student to pass.

In the end, students will not fail language classes. A major reason for this practice, as described in this section, is the cultural concept of collectivism. Now, I will turn to the second sociocultural theme identified in the focus group and semi-structured interviews – the perceived role of a language teacher in Uzbek society.

8.1.2 Perceptions of Teachers

In Uzbek society, teachers are perceived by the public as someone who has authority over learners due to their considerably higher level of knowledge and skill. A *language teacher*, specifically, is often “regarded as a master of language, and one who could interact with native speakers easily” (Feruza_10). Additionally, language teachers are recognized as cultural ambassadors (Klara_1) and are people who can bridge not only different languages but different cultures, as well (e.g., Uzbekistan and the United States). Thus, they are identified as being steeped into the lifestyle of the languages that they teach, and are often identified as language/culture experts.

These widely-recognized perceptions of language teachers mean there are high expectations for them to be effective in assisting students in the learning of language and culture (Klara_1). The assessment of students is reported by some participants as the relationship between the quality of a teacher and the learned knowledge of a student (Aziza_7 & Mohira_9). A teacher’s assessment practices then, are sometimes influenced by negative/positive results of

students' scores. When scores are negative, teachers often blame themselves for student failures, but when the students do well, teachers recognize the hard work they put into the course paid off.

8.1.2.1 Cognitions on Perceptions of Teachers and Assessment

There is a commonly held belief among the participants that if a student is unsuccessful, then the teacher is the person to blame. Often, the participating Uzbekistan EFL teachers blame themselves when students underperform. This tendency to blame themselves reveals that teachers sometimes hold the quality of their teaching accountable when students underperform on assessment measures.

Self-blame was discussed on three occasions during the focus group interviews. In Focus Group 1, Aisara explained that “if [student] results do not meet our requirements, or maybe [our own expectations] [then,] we should change [our] teaching process, [or] maybe [the] curriculum, [or the] book [that we use], or maybe the content of what we are teaching. I think so. It means that we [teachers] do something wrong, if we do not get a [good] result [that] we expect” (Teacher 1_FG1). This quotation emphasizes Aisara’s belief that there sometimes is a mismatch between what was taught and what was learned. Aisara, however, is focused on the course, and believes that if something had been changed about the course, then the student or students might have done better. She does not address other issues (e.g., the student did not study; the test was poorly constructed) that could have been detrimental contributing factors. Her belief was upheld by Madison from Focus Group 3, who added that “...bad results mean that we [have been doing] something wrong” (Teacher 4_FG3). Finally, Alsu, Chinara, and Dana from Focus Group 5 all agreed that if students perform poorly on assessments, then the teachers are to blame and something will need to change in their teaching practice.

Alsu: Yes, as a teacher, yes. As a teacher, if we have not [attained] the results that we expected usually, I think, it's my own opinion, that I think I did something wrong, what do you think-

Chinara: -Yes, exactly, an educated teacher should understand, should realize what she has done with the student

Dana: ...It is my fault

These three teachers all place the blame themselves for the lack of higher student performance on a test.

8.1.2.2 Perception of the Teacher Shaping Assessment Practices

The participating Uzbekistan EFL teachers are cognizant of how they perceive themselves and how they are perceived by students and the university community (e.g., other language teachers and administrators). Many participants reported that the EFL teacher is at fault when a student performs poorly in a course. The teacher might be insufficiently aware of who his or her students are, their interests, or perhaps the content of the course lacks some needed element that has been overlooked (e.g., more engaging instructional materials, more motivating teaching strategies).

Fuller awareness of how their students are helps teachers 1) determine an appropriate number of assignments for a course, which can influence how criteria for pass-fail assessments are established; and (2) write specific items and/or use distinct content for assessment tasks.

8.1.2.2.1 Student Interest

Student interest, which is believed by the participants to play an important role in learner motivation (thereby resulting in students doing well on assessments) plays a role in determining the number of assignments to feature within a course. The number of assignments has been reported by the participants to have an impact on how the breakdown of the continuous

assessment score is determined. An English language teacher from Andijan, Nodira, stressed the importance of knowing as much as possible about students' interests. Nodira reported that many of the language majors at her university are frustrated with the English writing class, because they are overloaded with too many assignments. Nodira explained how she determines an appropriate number of tasks based on her perception of student motivation.

Nodira: [Sometimes the curriculum suggests too many written papers, and students sometimes disagree. Then the students complain about how many writing tasks they must complete. For example, they should write journal entries, write two compositions, and create a portfolio, which includes critical reviews, of five, seven entries for the whole year. There are all different kinds of essays they do for the whole term, so sometimes they ...]

Interviewer: Do you ever decrease the number of writing assignments? Or increase the number of writing assignments?

Nodira: Sometimes I decrease. I see that students are overloaded and are not interested in the topics I want them to write.

Nodira's awareness of students' interests is useful information for her to determine an appropriate number of assignments for the English writing course. Seventeen Uzbek language teachers commented on how their students' interests motivated them to (1) write specific assessment items and/or use distinct content for assessment tasks; and, (2) help the teacher determine the number of assignments in a course, which can determine how criteria for pass-fail grades are established.

Students' interests can facilitate the creation of assessment tasks. Mohira is an English language teacher in a university in Kashkadarya which specializes in law. She explained that during her courses she tries to persuade students to study certain content matter by pointing out that it will one day be helpful in their professional careers. A week before each midterm examination, she supplies students with a list of the course topics that will be covered on the test. Most of the time she has at least

... one topic, [that the students do not like ... I try to explain to them the importance of the topic they do not like. For example, some students do not like famous court cases. If a student says in front of everyone that he does not like the topic, then, every student listens. The student will then tell me to do a different court case instead. So ... I will change court case used for the test] (Paraphrased for clarity, Mohira_9).

Mohira changed the topic that was to appear on the exam because she believed that if the students have an interest in a topic, then “that would be best motivation for them to do [well] on test” (Mohira_9). Mohira's comment echoed those of many Uzbek teachers who reported that they believe interest is a constructive motivating factor for learning language. Mohira, additionally, made an unanticipated connection to what Swain (1984) refers to as *bias for best*, which explains that language teachers (and test developers) who construct tests can assist students to successfully complete the test through different mechanisms (e.g., providing a glossary, including content that will intrinsically be interesting to students).

Overall, Uzbekistan EFL teachers reported they are concerned with how they are perceived by students, the community, and themselves. They explained that when students fail to pass their classes, they often blame themselves without identifying other circumstances that could have contributed to students' failures (e.g., poor test construction, ineffective study habits,

personal outside circumstances, etc.). To avoid negative perceptions on the EFL teachers and to have minimal failures of students, some participants expressed the need to pay attention to and understand students' interests. Uzbek EFL teachers reported that they try to understand their students' backgrounds and interests so they can better gauge course content and their own assessment practices.

I next turn to the third sociocultural theme identified in the focus-group and semi-structured interviews – Uzbekistan EFL teachers' ethical responsibility in assessment practices.

8.1.3 Ethical Responsibility

All participants felt that more effective EFL teachers have an ethical responsibility to the profession, which includes their students, school, and country. They reported that EFL teachers (and subject-area educators) are often confronted with ethical dilemmas around assessment matters, particularly in the reporting and communicating of accurate results. Klara commented that “[EFL teachers should be cognizant about his/her stance on such issues, and should also have strategies on how to deal with them appropriately]” (Paraphrased for clarity, Klara_1).

The ethical dilemma in communicating and reporting accurate assessment results stems from a place of good intentions by the Uzbekistan government. The Uzbek Ministry of Higher and Secondary Education supports the academic achievements of its students, and has tried to promote their hard-earned-academic efforts with financial scholarships. One language teacher explained that “For excellent marks, I think it’s a little higher than 500,000 so’ms (127 dollars), then, 400,000 so’ms (102 dollars), 300 (75 dollars) and so on” (Feruza_10). In particular, the scholarships are meant to support students from more distant regions (e.g., outside of Tashkent), so they would not need to work part-time jobs during university and will be better able to focus on their studies. The Ministry of Education understands the difficulties these students face and

has tried to level the playing field” (Diora_4). However, a few students sometimes put added pressure on themselves or their teachers to obtain higher grades. How an EFL teacher deals with such pressure from students reflects ethical considerations with both local and profession-wide repercussions.

8.1.3.1 Cognitions about Ethics and Assessment

Some participating Uzbekistan EFL teachers believe that student pressure on teachers (about money) is “not happening at all in Tashkent, but in the [rural areas], it still takes place, I know for a fact” (Klara_1). Those students who insist on pressuring teachers, and teachers who expect to receive money in exchange for giving students a higher score, are not using assessment practices to benefit the teaching and learning processes. In fact, teachers who decide to receive money from students in exchange for a higher score do not assess their students. Feruza explains that teachers “don't need to think about what grade to give students. You just count money, and [then] give them the grade which he [asked for]. Teachers who [receive] bribes, do not even [score] the papers. [If the student] asks for a B for example, the teacher will just give [him] a B. They will not [score] anything” (Feruza_10). However, there are Uzbek language teachers who say that they refuse such bribes because “It makes me feel very sad because that's a very obvious sign of corruption in our educational system, which is really bad. Really bad. Not only for educational system, but for the whole society” (Aziza_7). One language teacher claimed that she would absolutely not take any bribes, and she “would rather work in two, three, four places, but not live on bribery. Not at all” (Mukaddas_12).

The choice of an Uzbek language teacher to accept or reject a monetary offering by a student in exchange for a different grade is a sign of a dividing line between the language teacher who wants to adhere to an aspect of assessment ethics or one who rejects it. However, the

teachers in the semi-structured interviews all reported that every Uzbekistan EFL language teachers will have to make this decision one time or another in his/her career.

All of the participating teachers reported that bribery is part of an Uzbek language teacher's life. Mohira reported that bribery did not exist in the past because Uzbeks were part of a "more Islamic society before the USSR, and [Islam] does not allow bribery;" however, it was the USSR, which "introduced a lot of corruption [within] all spheres of Uzbekistan, including [its] educational system, so, beginning from entrance [exams], assessing, [to] assigning [grades]. Bribery has been the lead factor, so this was a bad influence. This all comes from the USSR regime" (Mohira_9). Bribery usually happens directly between a student and a teacher, but it can also happen among a group of students and a teacher. Diora explains what happened to her when English Language and Literature majors tried to convince her to give her a higher score:

Diora: They tried to convince me so that they can get higher marks...

Interviewer: Can you give me... a really... specific... example?

Diora: Offering, for example, "Is there anything we can do for you?" They don't directly say, "This is how much I'm going to pay you." They will go, "Is there anything we can help with you? Is there [...] anything you want? We can do it for you." Things like that.

Diora, like most of the Uzbek language teachers during the semi-structured interviews, became upset at the thought of bribery. We continued the conversation.

Interviewer: As a teacher, how does that make you feel?

Diora: Bad. Horrible.

Interviewer: What do you do to avoid that?

Diora: I'm usually very rude and I cut that off from the first beginning. I'll be like, "What am I doing here? Am I in business or am I trading something, or am I teaching you? Am I training you to do this in the future and is this going to be the same when you teach? How are you going to feel when you are treated like that when you are a teacher?" And they get the message.

Diora's decision to reject her students' bribe reveals an ethical decision, and suggests her more principled beliefs about what it means to be a responsible teacher. Diora believed that if she was able to help her students recognize the problem of bribery, then she would also be helping society move away from corruption. All teachers reported the current state of corruption in the Uzbek education system; however, they also reported that far fewer teachers currently solicit or accept bribes, and with each day moving forward, more teachers are opting out of the bribery system and only assign the grades students have earned.

Most teachers (if not all) during the focus groups and individual interviews reported that they believed their positive ethical responsibility for providing correct scores and reporting assessment results stems from their parents and how they taught them about the way to treat others and expect to be treated. All EFL teachers who participated in the semi-structured interviews drew a connection between what they learned from their parents and grandparents and how they report assessment results to students, colleagues, and administrators in an ethical way.

8.1.3.2 Assessment Responsibilities – Reporting Ethically

Svetlana, from Samarkand, talked about a time when she was trying to assess her students for a writing class for undergraduate English Language and Literature majors. Unfortunately, many of them did not pass her midterm examination because she felt that they had not studied sufficiently for it. When she shared the test results with two of her co-workers, they told her to curve the scores and make sure that all students should pass the test, no matter what. As Svetlana explained it, they said that she “would get in trouble, and it is the duty of an Uzbek teacher to be parental, to be kind to poor students” (Svetlana_2). Svetlana’s colleagues wanted the best for her and her students, but Svetlana thought differently. As she explained,

No matter [what] the consequences [are], [I feel it is important] to tell the truth, no matter who [you are talking to]. [For example], if it is the head of the department, [it is important] to talk straight and to be truthful. My mother told me to be truthful towards my profession because it's a sacred profession. My grandmother has the biggest influence on my life, I should say, because she was one of those first ladies in Uzbek communities who left home and studied at the university, way back [during the time of the] USSR [occupancy]. So, she had two bachelor's degrees and she was a director of high school all her life. She [taught] me to be truthful no matter the gender, if it's men or women. If you are right, if you are truthful, tell them. Don't be afraid. Don't be afraid.

Svetlana believed that she is responsible for reporting accurate assessment results, regardless of any potential social fallout. She felt that she was correct and principled in her assessment of students’ writing skills, and in the end, the administration “was happy to hear the truth. Lots of [changes] [needed] to happen” (Svetlana_2). Thus, from Svetlana’s point of view, how can students’ language learning progress if they are not working hard, studying hard, and using their

true test results to make themselves better language learners? Svetlana feels that teachers should not be afraid to confront students.

Like Svetlana, Umida credits her family, specifically her father, for her beliefs on how she interacts with students in principled ways:

My father is such [a sincere and a matter of fact type] of person. [He would have never said] “Oh, it's really a pity, [that a student has problems in his life].” [My father] never [said that]. [Instead], he always paid attention to the result. [He did not want to focus on the student’s background, or special problems]. If the result [of the test] is bad, okay. Nothing was changed.

Umida’s relationship with her father carried over into the relationship she builds with students. Umida reports that some students would characterize her as a strict teacher, but they also have high respect for her. Umida explains that she is able to maintain a “friendly relationship with them and they understand the consequences. If they don't study the subject that they are going to fail, I tell them from the beginning, and students are very smart” (Umida_5). Svetlana and Umida are examples of Uzbek university English language teachers who believe in the value of being ethically responsible with respect to their interactions with students, colleagues, and administrators and in the reporting of accurate assessment results.

8.1.3.3 Assessment Responsibilities – Scoring Ethically

Being an ethical assessor has been identified as important by some Uzbekistan EFL teachers. For instance, many of the participants commented that they try to be as fair and objective as possible when they score speaking tests. Feruza in Focus Group 4 explained that students often complain to the administration of the language department that “a teacher does not like me so she gave me a poor [grade]” (Teacher 6_FG4). The participants characterized this type of comment from

students as being avoidable. To score speaking tests ethically and responsibly, the teachers make sure there is an assistant in the testing room. This discussion sparked considerable interest from teachers in Focus Group 5, who talked about testing students' spoken skills, and how "students [often have] conflicts" and "to avoid conflict [we have] a second person" present (Teacher 5_FG5). Thus, during oral examinations there are at least two people in addition to the test taker in the testing room. The presence of two people is intended to demonstrate to the student that the scoring of the oral exam is unbiased and that they will receive a fair score. It is the case, however, that proctor teachers who serve on the testing commission could have had a student in a previous course (i.e., and therefore may have developed unexamined biases against some students). Also, the Uzbek teachers did not clarify whether the assistant in the room is also another scorer, or whether the assistant reviews or checks the teacher's score. Thus, the assistant's role – before, during, and after the assessment – remains unclear. What is certain is that there is another person in the room, and that the intention is to ensure fairness in assessment practices. Additionally, another reason for the presence of a second person seems like it is to avoid direct confrontations and verbal arguments during the speaking assessment interactions: with three people present, everyone is more likely to behave well; with two people, arguments could ensue. Because the scoring of a language learner's speaking abilities includes a considerable degree of subjectivity, I also asked if there was an assistant who was present during the scoring of writing. Focus Group 5 resoundingly remarked, "No!" (FG_5). In other words, only one teacher provided scores on each student's writing assessment.

8.1.4 Summary: Sociocultural Considerations

The first reported external factor is *sociocultural*, which is the relationship between social and cultural facets. The participants discussed three specific topics within a sociocultural perspective

that could shape their assessment responsibilities – collectivism, perceived role of a language teacher in Uzbek society, and ethical responsibility.

The first topic is the cultural value of collectivism, which addresses group relationships over individual pursuits. Because the participating Uzbekistan EFL teachers identified their society as collectivist, they viewed the classroom as an extension of what was commonly valued beyond the confines of the university school grounds. Most participants pointed to (and supported) the belief that *students should not fail*, because failing a student is directly contradictory to the goals of Uzbek society. The few EFL teachers who identified themselves as Russian-born, explained that they did fail students and were prepared for the social pressures or outcomes among their Uzbek-born counterparts. Thus, depending on the cultural group one self-identified as (i.e., ethnic-Uzbek, Russian-Uzbek), there might be a different understanding of what it means to fail students. In general, the participants reported they want the best for their students and they will try all that they can, in order not to fail them by 1) allotting more time for the student to finish all coursework and/or examinations; 2) changing the requirements on how the student will be assessed; or 3) only creating a final score for the student in the class (thus, producing an automatic pass).

The second topic identified by the participating Uzbekistan EFL teachers is the perceived role of a language teacher in Uzbek society. Language teachers are perceived to be steeped into the lifestyle of the languages that they teach, and are often identified as language/culture experts. This recognized perception of language teachers means there are high expectations to better assist students in the learning of language and, at the same time, culture. There is a commonly held belief among the participants that if a student fails, then the teacher is the one to blame. A

teacher's assessment practice, then, is sometimes influenced by negative/positive results of students' scores.

The third topic identified by the participants is an Uzbekistan EFL teacher's ethical responsibility. All participants felt that effective EFL teachers have an ethical responsibility to the language teaching profession, which also includes their students, school, and country. They reported that EFL teachers (and subject area educators) are often confronted with ethical dilemmas related to assessment, particularly in the reporting and communicating of accurate results. Often, Uzbekistan EFL teachers are confronted with students who try to bribe teachers to provide better scores in exchange for money and/or services. A few teachers reported that they believe this system of bribery stems from the USSR time in Uzbekistan, but now bribery is slowly withering away. Most (if not all) teachers during the focus groups and individual interviews reported that they believe their positive ethical responsibility for providing correct scores and reporting assessment results stems from their parents and how they taught them about the way to treat others and expect to be treated. All EFL teachers who participated in the semi-structured interviews drew a connection between what they learned from their parents and grandparents and how they report assessment results to students, colleagues, and administrators in an ethical way. In the next section I will address the second reported external factor, the sociopolitical factor, that the Uzbekistan EFL teachers identified during the focus group and semi-structured interviews.

8.2 Sociopolitical

The second external factor reported by Uzbekistan EFL teachers is *sociopolitical*, which is the interaction of the social and political systems of Uzbekistan. More specifically, the social system has been discussed above in Section 7.1, and here, I identify the political system as government

bodies or agencies at the local, national, and international levels. There are two main categories the participants reported as having a sociopolitical influence on their assessment practices: Presidential Decrees and Outside/Foreign Influence. I will discuss these broad categories in turn and provide relevant quotations from the participating teachers' focus groups and interviews.

8.2.1 Presidential Decrees

The first category is *Presidential Decrees* (PDs). These entail political documents written by first President Karimov and/or current president Mirziyoyev (see Chapter 3 for a Political History).

The most reported Presidential Decree (PD) by the participants is PD#1875 – the “Decree of the President of the Republic of Uzbekistan on Measures for Further Enhancement of the System of Teaching Foreign Language” (2012). This decree was written to systematically support all measures to enhance language learning and teaching of foreign languages in the Republic of Uzbekistan. The participating Uzbekistan EFL teachers discussed the decree positively. These teachers explained that the decree was successful in establishing governmental structures to provide services to influence, for the better, their language teaching, learning, and assessing practices (e.g., the in-service teacher education program at the Flying High Training Site). However, when the participants discussed the decree, they seemed unable to draw connections between such presidential decrees and their own assessment practices.

8.2.1.1 Cognitions about Presidential Decrees

The participants believe that PD#1875 is an important first step to enhance foreign language education in Uzbekistan. They credit its creation and implementation for some of the positive changes in their professional lives. Aziza explained, “According to our first [President’s] decree [PD#1875], and according to [the] decree of our current president, [PD# 2909], a lot of good things [are] happening now. [You] are here to teach us about teaching and testing, I think,

because of these decrees” (Aziza_7). Aziza’s comment references me as the researcher, because I was not only conducting research, but also teaching the participating Uzbekistan EFL teachers about principles of language assessment and testing.

The participants interpreted what may have been written within a decree as the reason or cause for their current assessment situation/context. For instance, Klara commented that “[There is much change in the way we teach and test language now at school, because of these presidential decrees] ... I am [directly] influenced at school, because, I need to know CEFR” (Klara_1). The participants report that the decrees are the underlying reason for their language curriculum and thus have helped to define the ways they need to assess their students and talk about assessment. In Focus Group One, the teachers explained the importance of PD#1875 and the connection to the CEFR:

Ajva: ... I think you heard about our presidential decree, 1875. Did you? Have you heard?

Interviewer: Hmmm... Please explain the presidential decree.

Klara: It is about our present development of teaching, learning English.

Interviewer: 120...209...2099?

Klara: No 1875. It was our first president decree.

Lena: In 2012.

Klara: And it influenced so much the educational system. We now, for example, I think in Uzbekistan's university we teach according to CEFR now. So, for example our students should have [attained] level B2 [by the time of their] graduation.

Klara cited the power of PD#1875 with a specific reference to the Common European Framework of Reference (CEFR). PD#1875 does not refer to the CEFR directly but states that “No later than before March 1st, 2013 [*sic*], ensure approval of new education standards describing detailed criteria for mastering foreign languages at each stage of learning” (PD1875).

8.2.1.2 Assessment Responsibilities Influenced

PD#1875 impacted Uzbekistan EFL teachers’ lives with the creation and establishment of new educational standards that they must know, be able to understand, and carry out appropriately.

The CEFR was adopted in December 2012, and Klara reported that “all university English teachers are expected to know the different descriptions and each level in the CEFR” (Klara_1). However, the CEFR is not the same as the National Standards (NS). The NS took most of its content from the CEFR (see Appendix N for the National Standards). Klara from Focus Group One also alluded to the importance of the CEFR when she explained that “our students should have [attained] level B2 [by the time of their] graduation” (Teacher 2_FG1). The decision to use the CEFR as a guiding language standard for teachers in Uzbekistan was an outcome of attaining the goal set forth in PD#1875.

Additionally, PD#1875 stated that “No later than before [*sic*] May 1st, 2013, ensure approval of new curricula and syllabus of ...higher education institutions.” To achieve this goal in the presidential decree, The Ministry of Secondary and Higher Education adopted the Pre-Service Teacher Education Training (PRESETT) curriculum in February 2013 for English Language and Literature majors. The Uzbek university language teachers who are teachers of English for technical and international specialties, however, do not yet have a set national language curriculum; instead, each university creates its own curriculum which has its own

assessment standards/practices, and “we are expected to follow what is in the curriculum” (Ulugbek_4).

PD#1875 is an external factor influencing university Uzbekistan English language teachers’ assessment practices because it impacts their lives with the establishment of the CEFR as the conceptual guiding standard and the development of the PRESETT curriculum for English Language and Literature or English Teaching majors. Universities that follow the PRESETT curriculum have authority figures – a rector (president), vice-rector (vice-president), and department head (department chair) – who make decisions about the curriculum’s testing procedures and how student outcomes should be reported. Participants reported that they are motivated to assess students on the existing curriculum because they want to comply with administrative policies.

When I asked the Uzbek teachers, “In your school, how large a role does the department head, the rector, vice rector, dean, have in the creation of your tests, your assessment procedures?” Svetlana responded, “They just regulate, and we have to make sure we are following the rules” (Svetlana_2). Mohira pointed out that “the department head plays the [largest] role in my assessment life” (Mohira_9), and has the most influence on how assessment practices are carried out at a university. “If I do not [comply with] her standard of testing, I am not doing well” (Mohira_9). The department head determines whether or not the curriculum is being carried out appropriately, and if students are being assessed fairly. Thus, “it is [in] our best interest to stick to the goals of the curriculum and also to test students according to the curriculum” (Nodira_6). There are two contextual factors of administrative authorities that contribute to an Uzbek teacher’s assessment practices: inspection and test development.

The most reported influence on a teacher's assessment practices was *inspection*, in which university authorities carefully examine and scrutinize policies and procedures. A common observation from Uzbek teachers was that administrators will hold language teachers accountable for assessments tied to the PRESETT curriculum. Ashura from Focus Group 2 states that the “authorities tried to [pressure us]. Yes, follow these curriculums. Even though each period they come and they check” (Teacher 1_FG2). Six Uzbek teachers who teach English Language and Literature majors, and who use the PRESETT curriculum, reported that one person from the administration would periodically come into their classroom and check to see if the teacher is adhering to the curriculum content. This inspection would happen unannounced: “[Before September all of the teachers came together and agreed on the curriculum that we would use and the lesson plans.] [However, that does not stop the administration from] interfering [in] our classes, [by] randomly [coming in and inspecting us and our lessons]” (Teacher 2_FG5).

Unlike the PRESETT teachers, no participant who teaches within an ESP curriculum reported having administrators enter into their classrooms. Instead, inspection by most ESP department heads happens as part of the semester's final examination period. Kamila explains that “We bring tests [to] our administration and they examine it closely...For example, [they will] review this test [to see] whether it is appropriate for the level [of students]” (Kamila_8). The Uzbek teachers who are part of an ESP curriculum at vocational and technical universities appear to have more autonomy in the development of their curriculum and the assessments that are a part of it; however, they are still inspected by the department head or a testing specialist when they have completed constructing their exams.

8.2.2 *Outside/Foreign Influence*

The second broad category of sociopolitical motivations is *outside/foreign influence*. These influences are operationalized as government agencies or related bodies and individual people from countries outside of Uzbekistan that shape the participants' assessment practices. The category of outside/foreign influence is closely related to Section 8.2.1, Presidential Decrees. The participants reported that the two countries which have had the most influence on language assessment practices in Uzbekistan are the United Kingdom (U.K.) and the United States of America (U.S.). Government-related agencies and specific people from these countries were welcomed to assist Uzbekistan in its PD#1875 language education reform which stated that Uzbekistan should "Get leading foreign education centres, international experts and specialists in related foreign languages involved in designing of education standards, curricula and syllabus, textbooks and organization of the education process" (PD, #1875).

Foreign entities assist Uzbekistan in (1) using the CEFR as a language standard to be used for language curriculum and language test development; (2) creating and implementing a language curriculum for English majors, and (3) providing Uzbek university language teachers with opportunities to interact with language teaching and testing specialists. These influences motivate the participants to learn how to (1) align tests to proficiency frameworks (e.g., CEFR), (2) use test specifications to develop items (question) and tasks, (3) employ techniques other than statistics (e.g., questionnaires, interviews, analyses of language) to gather information about the quality of language assessment, (4) apply rating scales to score speaking and writing performance, and (5) develop portfolio-based assessments. Svetlana commented that "teachers' jobs [are more difficult now] and have much [more assessment responsibilities than previously]. We need more help" (Svetlana_2). Svetlana's comment is a representation of most of the

participating teachers who reported their lack of confidence in meeting mandated assessment responsibilities. For example, Mukaddas, from Bukhara, would “like to see more assessment trainings at her school” (Mukkadas_12).

8.2.2.1 *The British Council*

The most frequently reported outside influence by the participants is the British Council. I will first explain briefly the British Council’s role in Uzbekistan, report on teachers’ cognitions around the British Council, and then, discuss participants’ explanations of how the British Council influenced their assessment practices.

8.2.2.1.1 British Council in Uzbekistan

The British Council is the United Kingdom’s well-known international organization for cultural relations and educational opportunities. It is a registered charity incorporated and governed by Royal Charter, though it is independent from the U.K. government. “The British Council works closely with all U.K. governments and devolved administrations to ensure strategic alignment to U.K. policy priorities” (BC website, retrieved on February 21, 2018). The British Council works closely with the Ministry of Higher and Secondary Education in Uzbekistan for the alignment of the CEFR and the creation of the PRESETT language curriculum. Currently, an ESP curriculum is being created by the British Council that is aligned to the CEFR, and will be intended for language teachers who do not teach English majors.

8.2.2.1.2 Cognitions about the British Council

The participating Uzbekistan EFL teachers believe that the British Council has supported Uzbekistan’s foreign language teaching and learning in a positive way. Most participating EFL teachers have a favorable attitude toward the British Council, particularly if the participating Uzbekistan EFL teachers are younger (in age and years’ experience teaching English). For

instance, Samia in Focus Group 4, who is a novice ESP teacher in Karakalpakistan, felt that “they [British Council] do a lot of things, especially bringing CEFR... they have a lot of trainings, a lot of things they have done that affects me, now they launched [an] ESP project. (Teacher 2_FG4). No teacher directly reported a negative or pessimistic attitude toward the British Council. Diora from Khorezm, who is a novice EFL teacher explains that she “appreciate[s] all they have done for Uzbekistan. Without them, we might still be in the Soviet system of education” (Diora_4). Diora’s comment shows that she believes there is a clear difference between the Soviet educational system and the new-founded Uzbek system of language education supported by the U.K.’s British Council. The language teachers who taught during the Soviet era did not say anything directly negative about the British Council or their organizational role in Uzbekistan. Instead, they have presented a neutral attitude toward them. For instance, Bibidana, who has taught for 33 years in Bukhara stated simply that “what they have done for us is okay. Lots of more work. We must follow what they have done for us” (Teacher 1_FG3).

8.2.2.1.3 British Council Shaping Uzbekistan EFL Teachers’ Assessment Practices

The British Council has influenced the language curricula that the Uzbekistan EFL teachers follow. Samia’s comment above addresses many of the things that the British Council has done that affect Uzbek language teachers directly. One of the largest factors is the implementation of the CEFR. Uzbek language teachers are now motivated to be able to (1) identify students’ language levels according to this widely employed framework and (2) align tests to the CEFR. The importance of the CEFR is discussed in Focus Group 1:

Albina: We teach according to CEFR now. So, for example our students should have [attained] level B2 [by the time of their] graduation.

Interviewer: So, for all universities, is B2 the level that you are expected?

Albina: In specialties [it is] C1 [but for others] [it] is B, B2.

Ajva: Is B, B2.

Albina: Yeah, and so it influences very much. So, we work hard, but there are some things which take time to improve. For example-

Leila: Challenges.

Albina: Challenges. For example, when students enter our university, we identify level and they should have level B1. But [they come to the school with] A2/B1, between A2 and B1. So sad.

Leila: B1.

Albina: B1, when coming to our university.

Leila; No, exactly B1. From Fergana: But, not all of them have B1 yet. We test that at [the] end [of their course of study]. We have to test that. It needs time.

Klara: [Students need at least five years to develop a higher level of proficiency.]

Topics exchanged among the teachers in the First Focus Group included discussion of practices of diagnostic and summative assessment. The participating teachers are motivated to conduct these types of assessments because they are required to help students exit the university with a specified level on the CEFR scale. Thus, they need to be able to design or align tests to the CEFR proficiency framework. The British Council revamped Uzbekistan's English Language and Literature English Teaching major curriculum in 2013 and titled it the *Pre-Service English*

Teacher Training (PRESETT) Curriculum. The aim of the new curriculum is to prepare future teachers of English to be competent in using the English language and in employing effective methods of language teaching. The participating teachers' assessment practices are tied directly to this curriculum because it provides assessment specifications for continuous, midterm, and final assessment measures. Svetlana explained that "Making tests based on test specifications is a first. The PRESETT curriculum lays it out but I am not given much guidance on [how to implement] it" (Svetlana_2). Creating tests from test specifications is a new challenge for Svetlana; before PRESETT she did not use test specifications, and many Uzbek language teachers are confused by what to do and how to do it. At the beginning of each semester of study, Umida emphasizes that teachers

make up our syllabus according to the PRESETT curriculum and all teachers [who teach the same course collaborate in the syllabus-creation process] ... Sometimes we design like this. Grammar teachers [work] together on [a shared syllabus and shared tests] for example, for first year students and others for the second-year students. (Umida_5).

Thus, sometimes, teachers work together to make tests from specifications because they have limited prior experience with this assessment practice.

PRESETT has a lasting impact on the way that the teachers assess their students, because the "PRESETT program pretty much changed the whole curriculum of our university, so we got rid of many subjects and we added many subjects like discourse analysis. We never had this at the school when I studied. I never studied this. British Council trained us to teach these subject with the CEFR. It gives us clear criteria to measure the levels of language" (Klara_1). Within the curriculum, language teachers are expected to use test specifications to develop items (test

questions) and tasks, and with classes that require spoken or written language, rating scales have been developed and are expected to be used.

8.2.2.2 Trainings from U.S.-Based Specialists

Trainings from U.S.-based Specialists are the second reported outside influence. The specialists are from three programs provided by the U.S. Department of State. First, Uzbek teachers discussed the experience of working alongside U.S. Peace Corps trainers. Second, they mentioned trainings with an English Language Specialist from the Office of English Language Programs. And finally, one language teacher reported practices she learned as a participant from an online assessment course co-sponsored by the U.S. Department of State and a U.S. university.

8.2.2.2.1 Cognitions Around U.S.-Based Specialists

The participating EFL teachers believe that U.S.-based specialists who come to Uzbekistan, or work with Uzbek teachers online, have the most up-to-date knowledge and skills in language teaching and language learning. They reported that they look forward to interacting with (and learning) from them about theory, research, and teaching practices. Ulugbek emphasized that he “like[s] the U.S. department of State because they provide [us] the most professional people. The English Language Fellows and Specialists have been positive role models as language teachers for our Uzbek language educators” (Ulugbek_2). Thus, when Ulugbek thinks of U.S.-based specialists, he thinks of English language teaching professionals.

Similar to Ulugbek, Mukaddas from Bukharra explains her positive experience interacting with an English Language Specialist from the U.S., who came to Uzbekistan to support language teacher development in conducting empirical research. She displays her favorable attitude toward U.S.-based Specialists in the following vignette:

I can remember it was a three-day intensive training on basics of research in applied linguistics. Although it was only a three-day training, it was very productive because we had a chance to learn much information related to types of research, research tools, data gathering and data analysis methods. We also learned the specific features of research conducted in the field of applied linguistics and what makes it different from research in other fields. [The Specialist] could prepare us for a real research which is conducted according to all the international standards. I can say this is my first experience of conducting a research in the field of applied linguistics according to the international standards and it was great. We planned our research – made a research design, identified the sample, developed the tools for gathering data and discussed all the phases of the research process. (Mukaddas_12)

Mukaddas' favorable attitude toward the U.S.-based Specialists is similar to Diliya's, from Focus Group 2. Diliya commented that "We like U.S. people and appreciate the information they bring us. They are not so serious as us Uzbeks are. They are very friendly and open and are very patient" (Teacher 4_FG2). Diliya and another member from Focus Group 2 were recently involved in a project with the U.S. Department of State that had just ended in Summer 2017. In the project they worked alongside an English Language Specialist on the development of an academic writing textbook. The participating Uzbekistan EFL teachers not only recognize professionalism in U.S.-based Specialists, but also have a positive attitude toward U.S. people's personality, as they are "open and are very patient" (Teacher 4_FG2).

The participants reported that their favorable attitude toward U.S.-based specialists did not waver when the government of Uzbekistan removed U.S. specialists from the country in

2005 (see Chapter 3). Klara was proud to report that she was (and continues to be) strengthening U.S.-Uzbek relations. She explained that,

When I read the book written by Mavlon Shukurzoda “U.S.-Uzbekistan Cooperation, Facts and Numbers” (2016), which is devoted to the 25th anniversary of diplomatic relations establishment between Uzbekistan and the United States of America, I can see many pictures of our university and even of myself in it. I feel that I am a part of this cooperation. I know I can do much more to contribute to the continuation of Uzbek-American mutual understanding in future and I welcome all U.S. citizens to continue building that relationship. (Klara_1)

Overall, the participating Uzbekistan EFL teachers have a propitious bent toward U.S.-based Specialists. They reported favorable beliefs and attitudes about the work that Specialists do and hope to continue building bridges of mutual understanding and relationships.

8.2.2.2.2 U.S. Based Specialists Shaping Uzbekistan EFL Teachers’ Assessment Practices

The U.S. based specialists reported by the participating Uzbekistan EFL teachers are U.S. Peace Corps trainers, English Language Specialists, and professors from an online assessment course in the U.S.

U.S. Peace Corps trainers influenced the assessment practices of Uzbek language teachers. This assessment practices include the creation of goals/objectives for lessons and the provision of feedback to students. Nodira worked as an Uzbek language educator charged with teaching Uzbek to Peace Corps volunteers in the late 1990s before she started teaching English at the university level. During her training experience to be an Uzbek language educator, she learned how to write goals and objectives for her language lessons and was taught the phrase, “be able to do...” (Nodira_6). She found this “American phrase absolutely intriguing,” because it

helped her understand that learning language “is more than knowing the grammar” (Nodira_6). Currently at her university, Nodira is responsible for conducting teacher training seminars for her colleagues. “[I think Peace Corps taught me everything that I use now in my trainings], and it gives a lot of positive results” (Nodira_6). She explains that “for example, I taught them, *by the end of the lesson, students will be able to* So, you should do action verbs” (Nodira_6). When working with both students and other teachers, Nodira emphasizes that goals are very important for language learning.

Umida worked as a Teaching English as a Foreign Language (TEFL) Coordinator for the Peace Corps. As a TEFL coordinator she worked closely with the Program and Training Unit comprised of American citizens and Uzbek nationals. She explained that all her language assessment knowledge is “from there” because she had “a lot of resources” and “you know, as a TEFL coordinator, I had a chance to analyze volunteers’ lessons, how they teach English at schools, and I had a chance to give feedback to them” (Umida_5). The way the lead American trainer provided feedback to the trainers in training and Peace Corps Volunteers influenced her greatly. Umida explained that she,

... tried to copy him, everything as he did it. [For example, he used] feedback sheets, and we would prepare everything [for him], and in the box [we brought] markers [and] papers for trainings. And [then we would] go to the trainings. He's an excellent trainer. He taught us how to make lesson plans, to give feedback. Then [at] the trainings ... By the language [he used], he taught us [everything that we should do], and he was both a good trainer and a good person (Umida_5).

Informed by her experiences with the trainer, Umida continues to use his techniques when she provides feedback to English language learners at her university, and she uses this experience

when serving as a language teacher educator. In ways comparable to Nodira's experiences, Umida has also received positive feedback from university Uzbek language teachers related to some of the lessons she learned while working with the Peace Corps TEFL Coordinator.

An English Language Specialist influenced the assessment practices of Uzbek language teachers, particularly with using statistics to analyze both test items and degrees of inter-rater reliability. Klara is an avid learner of English and she is also a new university-level English language teacher and worked with an English Language Specialist. Prior to teaching in settings of higher education, she worked in a middle school, with "students who were outstanding [in terms of oral language proficiency], sometimes better than teachers" (Klara_1). Her dedication to the English language profession motivated her to want to learn more about English language assessment. She approached me with two colleagues, and asked me to teach them how they can use basic statistics to make better-informed decisions about what they can do in their language classes. I tutored them in basic item statistics and how to calculate inter-rater reliability. "The calculation that you taught us, it was very helpful ... A lot of people today like to go on the computer, just put in numbers and get out a number, if you actually see how the calculation is created, there's a logic behind it. That's fascinating. I have used your methods on my exam scores from last semester. Wow. I wish I knew this information before" (Klara_1). Klara was analyzing the scores that she had previously assigned to students on a speaking test and she explained that "[before I thought that it was easy to assess speaking, but now I changed my mind. I now understand that achieving high validity and reliability is difficult to accomplish when scoring productive skills of speaking and writing]" (Paraphrased for clarity, Klara_1).

Specialists in language assessment from an online learning program sponsored by the U.S. Department of State had an influence on the assessment practices of Madison, from the

fourth focus group. She explained that the only thing she remembered was the four traditional criteria:

validity, practicality, what else? Reliability. The most important thing here [is] reliability. We have subjective assessment, and you know, it's my opinion, [but I don't think] that we [are] doing [it correctly]. Because [we have] practicality [issues], [as] it's sometimes difficult to administer [subjective tests], it [takes] a lot of time, it demands a lot of time. It requires too much time to check... It's not practical to [conduct subjective tests], and reliability [is an issue]

Madison's knowledge of language assessment and testing is based on the four traditional criteria. Furthermore, she recognized the difficulty of talking about these issues with colleagues in Uzbekistan because many of them "do not understand and don't need to" (Teacher 3_FG4). Madison believed that the knowledge of the "scientific ways to measure tests" should be the responsibility of the administration. She realizes that her rector and advisor "are trying to do anything [sic] that is possible to make changes" (Teacher 3_FG4).

8.2.3 Summary: Sociopolitical Considerations

The second reported external factor is sociopolitical, which is the interaction of the social and political systems of Uzbekistan. The participants discussed two specific topics within a sociopolitical perspective that could shape their assessment responsibilities – presidential decrees and outside/foreign influence.

The first category reported is Presidential Decrees. The most reported Presidential Decree (PD) by the participants is PD#1875 – the "Decree of the President of the Republic of Uzbekistan on Measures for Further Enhancement of the System of Teaching Foreign Language" (2012). This decree was written to systematically support all measures to enhance language

learning and teaching of foreign languages in the Republic of Uzbekistan. Although the participants' comments tended to reflect positive attitudes toward the presidential decree, they seemed unable to draw explicit connections between PD#1875 and their own assessment practices. The participants believe the decree impacted their lives for the better with the creation and establishment of new educational standards that the EFL teachers must know and be able to understand (i.e., National Standards, Common European Framework of Reference). Since the participants work within this national system, they are required to adhere to this framework by university authorities. These authority figures make most decisions regarding testing procedures and how student outcomes should be reported. The construction of a midterm or final examination, however, is determined by what was taken from the PRESETT curriculum (ESP teachers do not have a national curriculum) and pieced together by the language faculty. Participants reported that they are motivated to assess students on an existing curriculum because they want to comply with administrative policies. The Uzbekistan EFL teachers who are part of an ESP curriculum at vocational and technical universities appear to have more autonomy in the development of their curriculum and related assessment practices. However, the exams are also inspected by the department head or a testing specialist when their construction is completed.

The second reported category is outside/foreign influence. The participants reported that the two countries and organizations which have had the greatest influence on language assessment practices in Uzbekistan are the United Kingdom's British Council and the United States of America's State Department. Most participating EFL teachers have a favorable attitude toward both countries and their respective language-teaching-related organizations. They believe that specialists who come to Uzbekistan from these countries are very helpful and have the most recent knowledge and skills about language teaching (and language assessment).

8.3 Overarching Research Question

This chapter addressed the following overarching research question: What are the macro-environmental constraints and/or affordances in Uzbekistan that shape how EFL teachers provide meaningful assessment situations for their students? To address this inquiry, I asked the following three questions: (a) What are the macro-environmental factors Uzbekistan EFL university teachers report? (b) What are their cognitions surrounding those factors? (c) How do the reported factors and cognitions shape Uzbekistan EFL teachers' assessment practices? These questions address and identify the Uzbek language-teaching mind as “thinking as a function of place and time, through interaction and negotiation with *social and historical contexts* (Burns, Freeman, & Edwards, 2015, p. 589, [emphasis added]”). As identified above, the macro-environmental factors (i.e., sociocultural and sociopolitical) in combination with Uzbekistan EFL teachers' cognitions, influence their assessment responsibilities for their students. As evidenced by the data, we may conclude that over time, the macro-environmental facets are supporting and shaping meaningful assessment practices based upon:

- (1) The cultural value of collectivism;
- (2) The presidential decrees of Karimov and Mirziyoyev;
- (3) Societies' perceptions of language teachers;
- (4) The Uzbek ethical responsibility to selves and others; and
- (5) Influences of the British Council and the U.S. Department of State.

I will discuss how the participants referred to each macro-environmental facet as either a constraint or affordance on their assessment practices.

8.3.1 Constraints

The first macro-environmental constraint that shapes Uzbekistan EFL teachers' assessment practice is the cultural value of collectivism. This uniting value among the people of Uzbekistan limits, but does not completely remove, the power and assessment responsibilities of EFL teachers to make judgments concerning struggling students' language abilities. Laden within this value is the belief reported by the participants that all students, regardless of race, class, and ethnicity (e.g., Uzbek, Kazakh, etc.) should not fail the courses they take. Instead, students who are struggling with course content, classroom assignments, and/or examinations should be provided with sufficient opportunities and support so they can learn the required knowledge or possess the necessary skills to pass a course. In general, the participants want the best for their students. The participants will try all that they can including changing and improving their ways of teaching in order not to fail them, particularly, if there is a reason to better support Uzbek culture and way of life (e.g., marriage).

There are a few participants who reported that they go against the norm and decide not to pass students who have performed poorly in their courses. The participants who reported not following the cultural rule of not failing students in a course were brought up by Russian parents (and/or Uzbek-Russian), attended Russian-style schools, and/or continue to follow the Russian/Soviet way of education when they teach English. I will distinguish these EFL teachers from other participants and identify them as *Russophilic* – those who expressed having a positive bent toward Russian education and/or culture. The Russophilic participants emphasized that it is perfectly acceptable (and encouraged others) to fail students who do not deserve to pass (Nozliya_11). Additionally, the Russophilic participants who reported failing students were all older (in age) and taught languages for a number of years during the USSR time, where

assessment practices (and scores) relied on summative measurements only. The cultural view that students should not fail a course is a cultural constraint on every Uzbek EFL teacher.

The second macro-environmental constraint that shapes Uzbekistan EFL teachers' meaningful assessment practices are President Karimov and Mirziyoyev's Presidential decrees. The decrees were written to establish governmental structures to provide resources and services that influence, for the better, language teachers' teaching, learning, and assessing practices (e.g., PD#1875). The decrees constrain, or put limits on, the education system of Uzbekistan to help provide focus and direction for what should be taught, learned, and assessed. The participating Uzbekistan EFL teachers believe the decrees have impacted their lives for the better with the creation and establishment of new educational standards that the EFL teachers must know, be able to understand (i.e., National Standards, Common European Framework of Reference), and be able to apply to their teaching contexts. Most participants also noted that they appreciate the direction the country is moving, toward a more Western-style education system.

In contrast, the Uzbek university language teachers who are teachers of English for technical and international specialties do not yet have a set national language curriculum. Instead, each university program that features one of these specialties employs a locally generated curriculum guided by its own assessment standards and practices. These teachers have more autonomy on what to teach, how to teach, and how to assess. Many ESP teachers find teaching and assessing quite challenging, and some have reported they have to revert back to the time when they were learning languages during the Russian education system for their curriculum and/or assessment practices, or copy what the PRESETT curriculum offers. Although the Presidential decrees put constraints on the education system in Uzbekistan, they were also

created to change the existing Russian-style system for language education and adapt a more modern and/or updated approach to learning and teaching languages.

Overall, the cultural value of collectivism and President Karimov and Mirziyoyev's decrees are considered constraints by the participating teachers. These constraints help create meaningful assessment situations for students, because teachers are forced to become more aware of what they are teaching/assessing, why they are teaching/assessing certain facets of language, and how to better support student learning success.

8.3.2 Affordances

In this section I discuss the macro-environmental factors that are affordances for Uzbekistan EFL teachers' meaningful assessment practices.

The first macro-environmental affordance that influences Uzbekistan EFL teachers' assessment practice is societal perceptions of language teachers. An English language teacher in Uzbekistan is perceived as someone who is a master of the language and culture, and who has access to a worldwide network of academic, political, and business knowledge. Svetlana reported that her mother believed that because of a teacher's power, teaching "is a sacred profession" (Svetlana_2). Language educators are first and foremost teachers, who hold a certain level of power and control of the classroom, the content and information provided to the students, and how they test students. The relationship between the teacher's knowledge and skill base with the Uzbek cultural expectation of a teachers' role, presents a high degree of responsibility for the Uzbekistan EFL teacher.

The participating Uzbekistan EFL teachers' assessment practices are sometimes influenced by negative/positive results of students' scores. When scores are negative, teachers often blame themselves for student failures, but when the students do well, teachers recognize

that the hard work the teachers put into the course has paid off. Throughout the focus-group and semi-structured interviews, more EFL teachers reported looking inward and succumbed to self-blame when their students underperformed on assessment measures. Some also reported that they not only look inward but also outward at the class content and external measures and think about how those facets might be improved upon for better teaching and learning. As noted above, Aisara explained that “if [student] results do not meet our requirements, or maybe [our own expectations] [then,] we should change [our] teaching process, [or] maybe [the] curriculum, [or the] book [that we use], or maybe the content of what we are teaching. I think so....”

(Teacher 1_FG1). Changing these reported facets shows that Aisara is aware of the different affordances available to her for her assessment practices. She, therefore, uses the assessment results to make decisions about how to better the course. Because Uzbek teachers are respected for their knowledge and being authority figures, they have the ability to alter their courses to better support students’ language learning.

The second macro-environmental affordance is the ethics (moral code) of Uzbekistan EFL teachers. A teacher’s ethics serves to shape meaningful language assessment practices. During the focus groups and individual interviews, most teachers (if not all) reported that they believed their ethics (morality) is a result of parental influence. Family members taught them how to treat others and how they should expect to be treated in turn. The underlying lesson often repeated by the participating teachers is that others should be treated fairly and without bias, regardless of whatever the perceived outcomes of their actions might be. A common occurrence reported by all participants in the semi-structured interviews was that all Uzbekistan EFL teachers would be faced with an ethical dilemma – accepting or rejecting money in exchange for a higher score. As time moves forward, the participants reported that bribery has become a less

frequent ethical concern and is slowly being withdrawn from the education system. Nowadays, the reporting of accurate scores is becoming more common, and the grading system is becoming more transparent (e.g., the convention of including multiple raters on speaking tests).

The third macro-environmental affordance in Uzbekistan that supports shaping meaningful assessment conditions comes from the knowledge/skill base of outside/foreign entities. Foreign groups from the British Council and the U.S. Department of State provide multiple resources to support Uzbekistan EFL teachers' language assessment literacy. The participating teachers report that the knowledge base learned from teacher educators from the British Council and the U.S. Department of State includes outside influences that have not been constraining the assessment practices of Uzbekistan EFL teachers. Rather, they have been providing teachers with opportunities to explore different assessment practices that are meaningful and relevant to local contexts of language teaching.

Although most participants view foreign entities favorably, Svetlana commented that "teachers' jobs [are more difficult now] and have much [more assessment responsibilities than previously]. We need more help" (Svetlana_2). Her remarks (as explained above) emphasize that language teacher education regarding language assessment is a necessity for most Uzbekistan EFL teachers. All participants from the focus groups and semi-structured interviews agreed with Svetlana's ideas and commented positively that they would like to learn more information about language assessment. Yayra from Focus Group 5, who has (already) been to many different trainings on language assessment, explained,

As the teacher of English language I think all [foreign] teachers should have knowledge and skill of developing appropriate tests for their learners being familiar with guiding principles that govern good test design, development and analyses. I really want to be

aware with these important issues of language testing and I will search high and low for them because it is very important (Teacher 10_Focus Group 5).

Yayra's statement shows that she, like many of the participants, is willing to pursue additional professional development opportunities to become the best teacher (and assessor) of language that she can become. Additionally, some teachers, such as Diora from Khorezm, believe that the knowledge/skills base of language assessment that the foreign entities bring to Uzbekistan is part of a larger political process that will uplift Uzbekistan from the Russian education system. Diora emphasizes that "Without them, we might still be in the Soviet system of education" (Teacher 1_FG2). The majority of participating Uzbekistan EFL teachers value the knowledge and skill base of language assessment and would like to learn more in order to be able to employ more meaningful assessment practices with students.

8.4 Discussion

This chapter addresses the sociohistorical ontology in language teacher cognition research. Although the study does not address language teaching directly, it encapsulates the language assessment experiences of Uzbekistan EFL teachers that are situated within their assessment/testing practices and the sociocultural and sociopolitical environment (i.e., the macro-environment) of Uzbekistan.

Burns, Freeman, and Edwards (2015) emphasize that "the unit of research analysis in this ontological system placed emphasis on capturing thinking as a function of place and time operating through interaction or negotiation" (p. 592). Through teachers' self-reported discussions about different contextual facets that influence their assessment practices, I have been able to identify the different constraints and affordances from Uzbekistan's macro-environmental context that could shape the EFL teachers' meaningful assessment practices.

Additionally, through analysis, I identified a cultural and historical pattern in their thoughts, which was verbally expressed in their reported spoken language. The conceptual anchor for expressing their cognitions about sociocultural and sociopolitical aspects revolve around their personal relationship to Russia and/or Russia's influence in Uzbekistan's education system.

Depending on participants' background (e.g., family history, place of residence, type of higher educational institution, years of experience teaching English as a foreign language, etc.), some participants made a direct or indirect reference to the USSR. Their reference was reported in a positive and/or negative way, and thus, participants often compared life experiences either (1) while Russia was a significant force in Uzbekistan's educational system, (2) where the country is now in relationship to Russia, or (3) where they believe the country of Uzbekistan will be in the future and its role on the world stage. There are five overarching patterns identified in the teachers' reported cognitions and the sociocultural/sociopolitical macro-environmental facets in the chapter:

- (1) Russian-born/educated Uzbekistan EFL participants often did not comply with the culturally accepted rule of not failing students and were often ostracized from their Uzbek-ethnic born peers if they did not pass underperforming students.
- (2) Many of the participants consider the relationship between themselves and their students as a parent-child relationship and view the classroom as an extension of Uzbek society. Thus, they often blame themselves when students underperform. Those participants who are more Russian-influenced (e.g., Russophilic) emphasize that the language teachers' role is to be authoritative, and view the classroom as a separate and distinct entity from society.

- (3) Ethics (one's moral code) is directly connected to how a person was raised. All participating Uzbekistan EFL teachers reported that being fair to each person in society was important. However, some participants identified fairness in education differently and all semi-structured interview participants commented on corruption in the education system. The participating Uzbekistan EFL teachers also commented on that the cultural impact of corruption is waning slowly and will eventually be non-existent.
- (4) The Presidential decrees of Karimov and Mirziyoyev move the education system of Uzbekistan away from Russian style education, which directly influences the type of specialists who are coming to Uzbekistan to support curriculum creation, development, and integration.
- (5) U.K. and U.S. programs and people from the British Council and the U.S. Department of State introduce new, up-to-date, knowledge and skills about language assessment. All participants report that they appreciate the knowledge coming from such sources and would like to know more about contemporary principles of language assessment and testing.

As identified above, these five overarching themes that connect Russia's influence (or lack thereof) on teachers' cognitions correspond to Borg's (2003a) conceptualization of language teacher cognition. As explained in Chapter 2, Borg's (2003a) use of the term cognition reflects an integration of sources of knowledge, which includes schooling, professional coursework, contextual factors, and classroom practice. The categories listed above correspond to the category of contextual factors that influence L2 teacher cognitions.

When the participants reported on macro-environmental facets (i.e., contextual factors), such as Russia/USSR's influence, it was in a heightened affective state during the focus group and/or semi-structured interviews (e.g., during the discussion of ethics). The heightened affective state I am referring to was revealed more clearly during the semi-structured interviews but less so during focus group discussions. The former style of individual interviews seems to have provided more of an emotional safe space in which a participant could express his or her feelings more freely without the complication of feeling judged by other participating Uzbekistan EFL teachers. Sensitive issues (e.g., bribery, cultural mishaps, and politics) were discussed in detail during these interviews and the participants seemed to be expressing their opinions more openly. In fact, no participant commented they were unable to discuss certain issues from the questions I asked them. For future studies on teachers' cognitions that will be examined from a sociohistorical perspective, I recommend using interviews and other types of narrative inquiry such as journals over time. The written mode could provide a different type of space where language teachers might be open in expressing their opinions. One of the limitations of the present study is that I was unable to communicate with the Uzbekistan EFL teachers over time and in different modes (e.g., writing).

The discussions around Russia and/or the Soviet Union, specifically during the semi-structured interviews, allowed me to identify the participating Uzbekistan EFL teachers' emotional lived experience (Vygotsky, 1979). van der Veer and Valsiner (1994) quote Vygotsky, who stated, "The emotional experience arising from any situation or from any aspect of his environment determines what kind of influence this situation or this environment will have on the child" (p. 339). This quotation emphasizes that there is a dialectical relationship between the

environment (both micro and macro) and the emotions of people. Thus, in the study, the Uzbekistan EFL teachers reported experiencing a vortex of change (physically and emotional) at the national, local, and sociopolitical/sociocultural levels of their lives and have been expected to know new knowledge and skills about language, language teaching, language learning, and language assessment. With each day, the EFL teachers have become part of a new discourse on language teaching and assessment in Uzbekistan, with evolved cognitions and emotional lived experiences.

Now, I turn to Chapter 9, in which I will synthesize the results from the study (research questions one, two, and three) and show how Uzbekistan EFL teachers' assessment literacy is being constructed, renegotiated, and reconceptualized. The synthesis will show that researchers who wish to investigate the language assessment literacy needs of teachers should expand their research beyond the knowledge base of language assessment/testing to encompass the context, at both the institutional and larger macro sociocultural levels, that the language teacher is a part of. Language assessment literacy is not a static entity (e.g., having or not having literacy) but an evolving concept that fluctuates in every decision-making and action-taking purpose regarding a teacher's beliefs about assessment and the context that he/she is a part of.

9 CONCLUSION

In this chapter I demonstrate how Uzbekistan EFL teachers' cognitions on language assessment literacy are being constructed, negotiated, and conceptualized over time within evolving sociocultural and sociopolitical contexts. First, I locate the EFL teachers' cognitions on language assessment/testing practices within the setting of modern Uzbekistan (i.e., post 1991). Then I show the relationship between their maturing cognitions and how they identify, make sense of, and integrate assessment concepts into their everyday experiences.

Thereafter, I present opportunities and challenges for teacher educators to support the development of EFL teachers' language assessment literacy. In doing so, I see a relationship between the conceptualizations of and research into L2 teacher cognitions (see Chapter 2) in parallel to the models and practices of L2 teacher development. The relationship among the research generations, models of L2 teacher development, and the dissertation questions (with implications) are presented in Table 33.

Table 33 Research Generations, L2 Teacher Development, and Research Questions

Ontological Stances and Methodological Approaches	Models and Practices of L2 teacher development	Dissertation Research Question/Implications for SLTE
Individual	L2 Teacher Education in Infancy - Teacher Training	<p>RQ1: To what extent does the Language Assessment Literacy Survey (Kremmel & Harding, forthcoming), provide valid and actionable information about teachers' language assessment literacy</p> <p>Implication: Knowledge base of language assessment literacy; adaptable survey for all contexts for EFL teachers.</p>

Social	Bridging the Individualist to the Social Ontology: Teacher Development – Knowledge, Attitude, Skills, and Awareness (Freeman, 1989).	<p>RQ2: How do Uzbekistan EFL teachers talk about their assessment practices and justify the scores they provide for their students?</p> <p>Implication: What do they do and why? Looking at the micro-institutional (social) context), which informs L2 teacher educators how teachers choose to participate in assessment practices at the micro-institutional context.</p>
Sociohistorical	Re-conceptualizing the Knowledge Base of SLTE: Mindful Teacher Education	<p>RQ3: What are the macro-environmental constraints and/or affordances in Uzbekistan that could shape how EFL teachers provide meaningful assessment situations for their students?</p> <p>Implication: How they participate in macro-environmental contexts.</p>

The results from each research question provide specific information that, taken together, would support a well-rounded L2 teacher education program. I will take the phrase “models and practices of L2 teacher development” and call it second language teacher education (SLTE). Freeman and Johnson (1998) explain that SLTE is the formal label that describes the sum of various interventions used to develop professional knowledge among practitioners, which includes pre- and in-service L2 teacher education. I want to show how each research question and its results build upon each other. (I do not advocate for a specific model of second language education but see each research generation in L2 teacher cognition research supporting L2 teacher development. The results of each research question provide specific information that will

support L2 teacher educators when working with language teachers on language assessment practices.) Finally, I discuss future directions for research and articulate how the dissertation contributes and extends the academic fields of L2 teacher cognition and language assessment literacy.

9.1 Situated EFL Teachers' Cognitions in Uzbekistan

The colonial era of Uzbekistan ended in 1991 with the fall of the Soviet Union, which affected the cognitions of Uzbeks in different ways. Some believed the removal of the Soviet Union would bring back Uzbekistan's traditional cultural values and education system. However, some thought the Soviet system of government, education, and business that had been established in Uzbekistan was superior, and they preferred not to see the removal of the USSR (personal communication, Azizov, July 2017). Thus, there are conflicting cognitions and emotions surrounding the Soviet Union's involvement in Uzbekistan.

Although twenty-seven years have passed since the period of initial Uzbek Independence, colonialism's impact is still deep in the minds and hearts of the citizens of Uzbekistan. The impact has surfaced in the Uzbekistan EFL teachers' reported cognitions throughout the dissertation study. Even though there was only one participant out of fifty-three who explicitly seemed hostile towards the Russian/Soviet Union's effect on the Uzbekistan educational system (see Chapter 8), most teachers did not report antagonism. Instead, participants reported both positive and negative emotions when they referred to different periods in their lives, made connections between pre- and post-1991 independence, and discussed the future of English language education in Uzbekistan with or without the involvement of the Soviet Union or modern day Russian. The inclusion of time as a variable (i.e., 1991) helped me identify how Uzbekistan EFL teachers' cognitions shape their assessment practices.

When the participants reported their cognitions – through surveys, focus group interviews, or semi-structured interviews – three overarching voices emerged from the data: professional voices, generational voices, and cultural voices. The results of the first research question illustrated the participating Uzbekistan EFL teachers' professional voices. The results of the second research question brought out generational voices (i.e., those teachers who are younger and older in age). I identified distinct cultural voices (i.e., traditionalist and Russophilic) from the results of research question three. I will briefly discuss the results of each research question and the voice(s) that emerged from each. Then, I address how each aspect has implications for SLTE. I then show how the results from the three research questions could build upon each other to create a meaningful teacher education program on language assessment literacy for the participating Uzbekistan EFL teachers.

9.2 The Professional Voice: Identifying Academic Assessment Concepts

Results emerging from the first research question were intended to serve as a window into the language assessment literacy of Uzbekistan EFL teachers. The Language Assessment Literacy survey contained items that addressed specific skills and knowledge about assessment/testing practices that professionals and academics from the language testing community find to be important (see Chapter 2 for a detailed discussion of the survey).

The overarching questions in the survey guided the participants to think about language assessment literacy from a language teachers' perspective. More specifically, the survey asked two questions: (1) How knowledgeable do people in your chosen group/profession need to be about each aspect of language assessment; and (2) How skilled do people in your chosen group/profession need to be about each aspect of language assessment? The participants

responded to each survey item as if those items were devoid of cultural nuances. In other words, no item specifically said, “In Uzbekistan...”

The outcomes of the first research question exemplified a homogenous population of participating Uzbekistan EFL teachers, and showed a collective professional EFL teacher voice. Their overall responses supported Taylor’s (2013) initial premise that classroom teachers believe they should have more knowledge/skills (related to assessment) with language pedagogy. Additionally, the comparison of factor analysis results (mine and theirs) supported Kremmel and Harding’s survey instrument as a valuable tool that could be used in a variety of specific contexts (i.e., Uzbekistan) to find useful information for teacher educators to inform practice.

9.2.1 Implications for Language Assessment Literacy Teacher Education

The Language Assessment Literacy Survey addressed one way to identify classroom language teachers’ language assessment literacy needs. How a teacher educator ultimately responds to and uses information gained from teachers and trainees depends on the context and aims of the professional teacher education program. Some programs focus more on what has come to be known as teacher *development*, while others emphasize teacher *training*. *Teacher development* is a term used in the literature to describe a “process of continual, experiential, and attitudinal growth of teachers... some of which is generated in preprofessional and professional in-service programs” (Lange, 1990, p. 318). Teacher development is intended to continue over the entire professional span of an educator’s career. *Training*, on the other hand, focuses on teacher knowledge of the topic to be taught, and of the methodology to teach it. Training is particularly different from development because it is typically shorter term and less comprehensive in nature.

An important focus of a traditional SLTE program within a training model is to make sure that knowledge and skills of teaching are distributed to teachers. There is a commonly held

belief among language teacher educators that in this transfer of information, one will be able to produce better, more qualified language teachers, and subsequently, better language students (Kumaravadivelu, 1994; Wright, 2010). For a teacher educator who only focuses on the knowledge base of language assessment, the responsibility of the educator would be to distribute necessary knowledge and skills that would support teachers' assessment practices. This pre-Freirean information-transmission principle/banking model (Freire, 1970) is conceptually limited, in contrast to a teacher development model. Similar to the commonly held belief above, teacher educators hope that the responsible assessment practices a language teacher learns will lead to superior decisions being made about their students' language abilities.

I discuss two different ways teacher educators operating in a traditional SLTE program (e.g., training model) would be able to use the results of the Kremmel and Harding survey instrument. First, teacher educators can use the results of each survey item and identify which ones the participants think are most or least important. If teacher educators in Uzbekistan only used the item-by-item survey results, then the trainers would obtain a general understanding of the assessment needs of the EFL teachers. All items were viewed as important (or somewhat important) to a classroom language teacher's profession. For teacher educators, this result can be daunting and non-informative because everything is viewed as important. The teacher educators would need to be very well versed in all aspects of language assessment and testing and be able to explain, answer questions, and discuss all aspects of the knowledge/skill base. There are many language testing and assessment textbooks available (e.g., Bachman & Dambök, 2018; Bailey & Curtis, 2015) that could be used to aid a teacher educator to disseminate this information to language teachers.

Second, the results from the survey can help teacher educators identify the type of stakeholder(s) who are present in their SLTE program. The Language Assessment Literacy survey (Kremmel & Harding, forthcoming) was intended to be used with multiple stakeholders (i.e., testing professionals, classroom teachers, administrators). Even though most program participants in an SLTE program would self-identify as classroom language teachers, some could also have program evaluation or other administrative responsibilities. A teacher educator who only focuses on disseminating information could use the results from the Language Assessment Literacy Survey to gauge the syllabus content, tasks, and assignments, to address each specific population. The challenge for the teacher educator would be to learn how to tailor language assessment course content to the needs of the stakeholder group in the SLTE program.

Using the Language Assessment Literacy Survey – and a similar survey – would provide teacher educators a glance into the assessment and testing needs of a desired population. However, teachers' cognitions about the assessment items from the survey can change over time. To understand teachers' beliefs as both shifting and contextualized, Burns, Freeman, and Edwards (2015) noted that "Teachers' (re)conceptualization and (re)construction of their experiences, previous knowledge, and personal beliefs were seen to respond to 'both macro- and micro-level contextual factors in their classroom, schools, and communities'" (p. 590). Although the knowledge/skill base of assessment is important to know, and the Language Assessment Literacy survey helps teacher educators identify the needs of teachers, I believe we should move beyond the information-transmission model and take into consideration the sociocultural contexts in which the teaching, learning, and assessing take place. This information would provide teacher educators more information to help support teachers' language assessment literacy development as opposed to training only. I do want to note that the Language

Assessment Literacy survey is a good place to start developing a training model of teacher education or a developmental model. Unfortunately, some teacher educators are limited in time (e.g., the amount of interactions with teachers), training opportunities, and resources to support language teacher development of assessment literacy, and would only be able to focus on distributing the knowledge and skills of assessment to language teachers.

9.3 Generational Voices – Making Sense of Academic Assessment Concepts

The second research question addressed the participating Uzbekistan EFL teachers' contexts of practice. I identified a pattern in the teachers' reported cognitions and found a difference in how the younger and older participants view Assessment-*of*-Learning (i.e., Pre/During Learning) and Assessment-*for*-Learning (i.e., Post Learning) practices.

The younger EFL teachers reported they find doing Assessment-*for*-Learning (Pre/During Learning) to be a valuable, worthwhile, and necessary endeavor. These teachers reported being more open to trying out different assessment practices, which could support them being more effective facilitators of student language learning. However, the older (and more experienced) EFL teachers tended to have a negative attitude towards Assessment-*for*-Learning (Pre/During Learning). The discrepancy between the younger and older participants can be identified in their reported cognitions. For instance, teachers' experience teaching during Soviet-Era Uzbekistan versus post-USSR; being raised pre- or post-Independence (i.e., 1991); and/or attending an Uzbek, Russian/Soviet, or mixed-style secondary school all played a role in whether the participating teachers attempted to do Assessment-*for*-Learning (Pre/During Learning). Those teachers who believe more in the value of the Russian/Soviet style education system did not appear open to trying Assessment-*for*-Learning (Pre/During Learning) ideas. They seemed to value Assessment-*of*-Learning (Post Learning) practices more highly.

There was less of a discrepancy between the younger and older participants with respect to *Assessment-of-Learning* (Post Learning). Universities in Uzbekistan have tended to emphasize test results over classroom based practices/assessments. Many of the participating Uzbekistan EFL teachers who grew up and/or taught in a Soviet/Russian and/or Uzbek style school reported that conducting classroom-based formative assessments is a new concept. However, all teachers reported they took part in *Assessment-of-Learning* (Post Learning) testing practices at their local universities. Even though the younger and older participants took part in the midterm and final test creation process, the results of research question two revealed most teachers do not feel comfortable designing any type of test (i.e., midterm and final) because they often felt that they lack requisite knowledge and skills of language testing. These teachers questioned their test development abilities. Many participants reported that the test development processes at their local teaching contexts are unfair and not valid measurements of what is being taught in their classrooms.

Although the participants reported doing *Assessment-for-Learning* (Pre/During Learning) and *Assessment-of-Learning* (Post Learning), they generally did not score students using standard assessment/testing techniques for the betterment of student's learning. However, such was the intention of many of the participants. There was a discrepancy between what the participants reported they did, what they thought about what they did, and how they explained how they assigned scores to students. The younger and older participants reported similar procedures for how they kept track of student progress and eventually provided each student with a score, regardless of the teacher's reported cognitions. Most participants assign students a score based on a general feeling rather than on empirical data. These results suggest that there is a relationship between the participating Uzbekistan teachers' knowledge/skill base of assessment

(i.e., survey results), attitude towards it, and awareness (or lack thereof) of how to implement Assessment-*for*-Learning (Pre/During Learning) and Assessment-*of*-Learning (Post Learning) practices.

9.3.1 Implications for Language Assessment Literacy Teacher Education

The results from the second research question suggest that a teacher educator needs to understand more than the knowledge base of L2 assessment/testing to better support the assessment literacy development of classroom teachers. Teacher education is becoming closer to a constructivist model of education. For instance, Freeman (1989) offers a closer examination of language teacher education, both its subject matter and its processes. He created a model of four constituents that interact through the teacher's decision making, which are made up of knowledge, attitudes, skills, and awareness, also known colloquially as the KASA model (see Figure 20). This descriptive model supersedes the traditional knowledge-transmission model of teacher education that only addresses knowledge and skills.

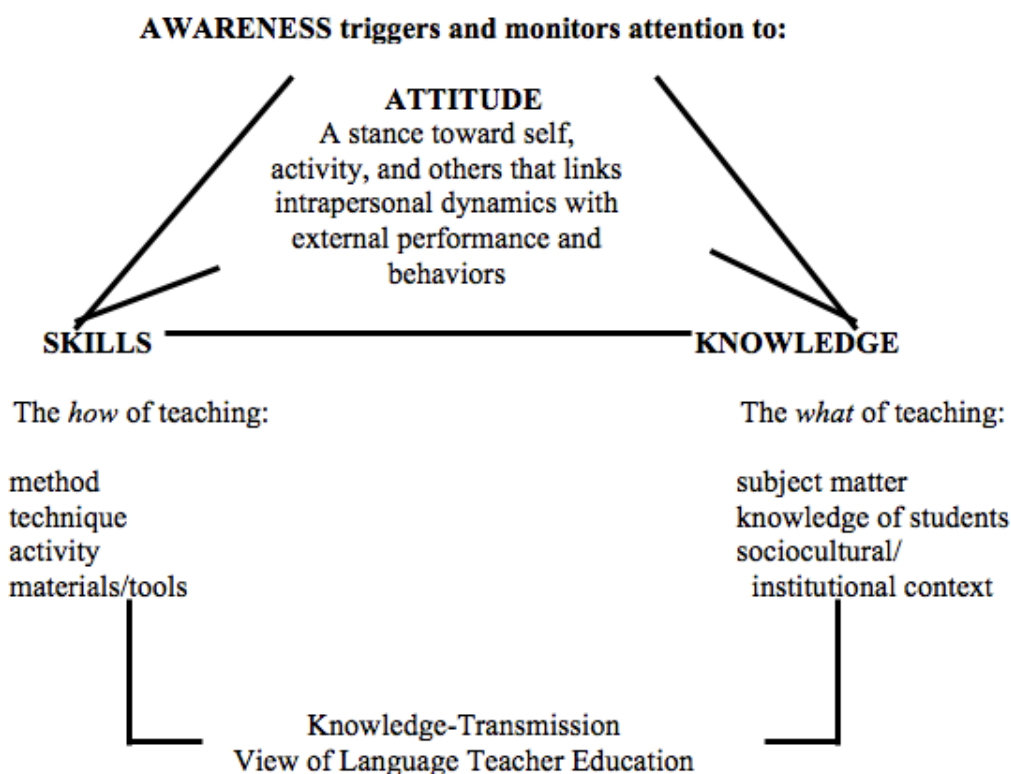


Figure 20 KASA Model

Two facets of the model, awareness and attitude, draw attention to the process of teacher education. The awareness component of Freeman’s model is important because, “much of what happens in teaching is unknown to the teacher” (Richards & Lockhart, 1994, pp. 3-4). The language teacher educator should support the language teacher to move from states of being unaware to states of being aware by helping them “activate, notice, select, direct, and so on by raising questions, providing observational data, and discussing the teacher’s concerns with him or her” (Bailey, 2006, p. 37). Additionally, Freeman’s model contains *attitudes*, which he defined as “a stance toward self, activity, and others that links intrapersonal dynamics with external performance and behaviors” (1989, p. 36). This definition is similar to the definition that is predominant in the social ontological research tradition (see Chapter 2) in language teacher

cognition research, in which this period placed concern about the “processes of learning to teach” (Burns, Freeman, & Edwards, 2015, p. 59) and paid particular attention to how the surrounding contexts, both internal and external to the person, shapes thinking.

Freeman’s (1989) KASA model, which is a constructivist model of teacher education, presents a supportive framework for teacher educators who will work with Uzbekistan EFL teachers on developing their language assessment literacy. From the results of research questions one and two, I found that Uzbekistan EFL teachers conduct *Assessment-for-Learning* (Pre/During Learning) and *Assessment-of-Learning* (Post Learning) practices. However, the participants reported they are unable to make sense of how to put these aspects of assessment into practice. The key concept a teacher educator needs to keep in mind for these teachers is awareness. Teacher educators in Uzbekistan working with these teachers would need to be cognizant of the two different populations of classroom teachers: (1) younger teachers who generally have been exposed to and have a positive attitude towards different assessment practices; and (2) older teachers who have negative attitudes about and less exposure to varying assessment practices. The teacher educator’s challenge then, is to bring awareness to both groups of people, and at the same time, support all teachers to have a positive attitude so that teacher change can happen. A teacher educator could incorporate techniques such as diary entries, observations, assessment monthly meetings, and the use of case studies to interact with the language teacher in the written or verbal mode. Also, a teacher educator could incorporate sociocultural models (e.g., Vygotskian or post-Vygotskian constructivist models) that include the role of mentors or facilitators in expanding teachers’ zone of proximal development (ZPD). Vygotsky’s ZPD metaphor states, “social interaction actually produces new, elaborate, advanced psychological processes that are unavailable to the organisms working in isolation” (Vygotsky

1989, p. 61). Thus, a teacher educator could use social interaction to support the development of a teachers' language assessment literacy.

9.4 Cultural Voices: Integrating Academic Concepts

The third research question addressed the sociocultural and sociohistorical contexts in which the participating Uzbekistan EFL teachers' assessment practices are situated. Specifically, I examined the macro-environmental constraints and affordances reported by the teachers, which spanned space and time. I was able to identify two-types of voices from the teachers' responses: traditionalist and Russophilic.

The participating EFL teachers whom I identified as having a traditionalist voice adhered to conventional Uzbek cultural values. These values were identified by me as both possibilities and barriers to integrating assessment concepts into their everyday teaching and assessing practices. Traditionally Uzbeks identify themselves as a part of a collectivist society, which means they value group relationships over individual pursuits. The participants want the best for students and they will try all that they can in order not to fail them. Many of the traditional-voiced participants consider the relationship between themselves and their students as something along the lines of a parent-child relationship, and thus view the classroom as an extension of Uzbek society.

The traditionalist voice contrasts with participants who portrayed a more Russophilic voice. Some of these participants were Russian/Soviet-born/educated and did not directly identify with the traditional values of Uzbekistan as explained above. Instead, they identified with the Russian and/or Soviet style of education. For instance, Russophilic participants often did not comply with the culturally accepted rule of not failing students and were ostracized from their Uzbek-ethnic born peers if they did not pass underperforming students. Those participants

who are more Russian/Soviet-influenced emphasize that the language teachers' role is to be authoritative, and view the classroom as a separate and distinct entity from society. This view contrasts sharply with participants who held a more traditional Uzbek perspective on the teacher-student relationship.

Both traditionalist and Russophilic participants noted they appreciate the direction the country is moving toward with a more contemporary education system, particularly in the field of foreign language learning. This more recent education system contrasts with the rote memorization, grammar-oriented learning of languages that they were previously accustomed to. The participants appreciated the U.K. and U.S. SLTE programs that introduce more recent, up-to-date, research-based knowledge and skills about language assessment. (Updated methods do not disregard rote memorization, as it could be useful in some circumstances.) All participants reported they appreciate the knowledge coming from such sources and would like to know more about contemporary principles of language assessment and testing.

9.4.1 Implications for Language Assessment Literacy Teacher Education

The results from the third research question suggest teacher educators should have deeper appreciation of the macro-environmental contexts within which prospective and practicing Uzbek language teachers operate. These factors could serve to facilitate or impede learning. The challenge for a teacher educator is to reposition his or her cognitive focus, from a perception of teaching and assessing from a behavioral to a constructivist view to bring together all aspects of a language teacher's life. For "teacher educators have come to recognize that teachers are not empty vessels waiting to be filled with theoretical and pedagogical skills; they are individuals who enter teacher education programs with prior experiences, personal values, and beliefs that inform their knowledge about teaching and shape what they do in their classrooms" (Freeman &

Johnson, 1998, p. 401). Although Freeman and Johnson's (1998) work continues to be the anchor of thought for the field of SLTE (Lee, Murphy, & Baker, 2015), there is a growing perspective that has gradually begun to wield more influence in second language teacher education. It is called *mindful L2 teacher education* – which is both a practice of L2 teacher education and a method of research.

Mindful L2 teacher education, a term identified by Johnson and Golombek (2016), is about engaging in dialogic interactions that can help assist teachers as they transform their knowledge, disposition, and skills through responsive mediation. In other words, responsive mediation happens inside the practices of teacher education as it unfolds, which looks at the teacher as a whole person and thus “requires establishing a sense of a teacher's [emotional lived experience], both past and present, as well as recognizing teacher educators' own complex interplay of cognition and emotion, originating in and reshaped through their [emotional lived experience]” (Johnson & Golombek, 2016, pp. 168-169). From this understanding of a teacher's emotional lived experience, then, responsive mediation identifies challenges, tensions, and excitement of language teachers (Johnson & Golombek, 2016). The identification of these moments may represent potential *growth points* (Johnson & Golombek, 2016; McNeill, 2005) – a moment or series of moments when teachers' cognitive/emotional dissonance comes into being. Johnson and Golombek (2016) argue that the “responsive mediation directed at the growth point creates the potential for productive teacher learning and development” (p. 45). Teacher educators thus need to forge *intermental development zones* (IDZs) – multiple and sustained opportunities to think together about the problems, concerns, and/or issues at hand. The IDZ process also involves the offering of assistance that is responsive to immediate needs and emerging capabilities (Johnson & Golombek, 2016, pp. 168-169). It is thus within the IDZ where discussion ensues between the

teacher educator and the language teacher so that potential growth points may be targeted and responsive mediation can occur. The sociocultural model of responsive mediation is more of a teacher developmental model than a training model (as explained in 9.2.1).

In Uzbekistan, the challenge for teacher educators is to forge these intermental developmental zones among EFL teachers who have different cultural voices (i.e., traditionalist voice and Russophilic voice), generational voices (older and younger), but, a similar professional voice. The teacher educator must understand that each individual teacher has his or her own concerns, frustrations, excitements, and experiences. The life histories and experiences of language teachers differ in both major and minor ways. Though there may sometimes be similarities between them, no two teachers' teaching and assessing experiences are completely the same. Thus, the traditional approach to teacher education no longer suffices, and we cannot rely solely on the distribution of assessment knowledge and skills from a traditional (e.g., training model) SLTE approach to truly enhance language assessment literacy of teachers. Teacher educators should not also simply integrate reflective techniques in the hope we can alter teachers' attitudes and awareness about different *Assessment-for-Learning* and *Assessment-of-Learning* practices. Rather, teacher educators charged with working with pre- and in-service language teachers should understand that they are active participants in teachers' developments. (See Appendix O for brief outline of an in-service teacher education program on developing assessment literacy for Uzbekistan EFL teachers.) L2 teacher development is not linear (Kiss, 2012) and it is the teacher educator's job to create the social conditions necessary for development to occur, which is in the engagement of the dialogic interactions.

9.5 Future Directions

The dissertation's findings, discussions, and implications are based on the participating Uzbekistan EFL teachers reported assessment practices, thoughts, and ways that they score their students with Assessment-*for*-Learning (Pre/During Learning) and Assessment-*of*-Learning (Post Learning) measures. Based on the results, there are two future studies I would like to pursue: first, a teacher cognition study that also features a classroom observation component of Uzbekistan EFL teacher(s) in action with student interviews; and second, another teacher cognition study that includes a discourse analysis component of post-observation conferences between teacher educators and Uzbekistan EFL teachers.

First, an observation component to a teacher cognition study on assessment literacy is necessary. To confirm that what the participating Uzbekistan EFL teachers reported and discussed throughout the report feature in the current dissertation holds true, I will eventually need to move beyond data collection procedures of survey, focus group, and semi-structured interviews to classroom observations. When that data is collected, analyzed, interpreted and combined with the current study's findings, I will be better positioned to more fully discuss the construct of intentionality, which was briefly discussed in Chapter 2 as a future direction for the L2 Teacher Cognition literature. This construct is used by L2 teacher cognition scholars to research the inner lives of teachers in action (e.g., in the classroom). Additionally, not only would the research direction be expanded to include a classroom observation component, but I would also extend the study to include student voices. Then I would be able to compare what the EFL teachers do with assessment with the students' own perceptions and beliefs about whether the teachers' assessment practices are meaningful for them.

Second, a discourse analysis study of a teacher educator in conversation with an Uzbekistan EFL teacher after a post-observation conference would support Taylor's (2013) call for research to examine "language and discourse when engaging with a non-specialist [in language testing] audience" (p. 406). This study could look at the sociocultural domain of the teacher educator-teacher interaction in post-observation conferences and examine the language that is used by the teacher educator to promote teachers' assessment literacy development. Additionally, this future research direction would use the methodology of mindful teacher education (see above), which is about engaging in dialogic interactions that can help assist teachers as they transform knowledge, disposition, and skills for themselves through responsive mediation. Not only would such a study support research in the academic field of L2 Teacher Cognition, but it would also inform the Language Assessment academic community on how to interact with non-specialists in language testing. This future study could take place as part of a language assessment practicum, or as a model for the quality of interactions that could be planned as a part of such a practicum, or as part of formative conversations between an assessment specialist mentor and a classroom teacher.

9.6 Contributions to Academia

The dissertation is a contribution to efforts within Applied Linguistics generally and L2 teacher cognition and language testing in particular. The research methodology and content matter of the dissertation contribute to the growth of language teacher cognition research and the academic field of language assessment literacy.

9.6.1 L2 Teacher Cognition

In the present study I expanded the L2 teacher cognition research agenda by featuring processes of language assessment as its subject matter. The first research question continued the tradition

of the individualist ontological research tradition, while the second addressed the social research tradition. The third research question expanded the analysis of the relationship between L2 teacher cognitions and language assessment literacy to include the sociohistorical ontology research generation. To my knowledge, there are no studies to date that have examined the relationship among L2 teacher cognitions, language assessment literacy, and macro-environmental factors from a sociocultural and sociohistorical perspective. In doing so, I have applied a more eclectic view of the language teacher mind, which reflects Borg's (2006) view of L2 teacher cognition as an integration of knowledge sources that includes schooling, professional coursework, contextual factors, and classroom practices. Additionally, I explored how the participants' assessment and testing knowledge, beliefs, and feelings shape the assessment practices they use, which has not been researched in the L2 teacher cognition research agenda.

9.6.2 Bridging L2 Teacher Cognition and Language Assessment Literacy

In conducting the study, I aspired to bridge the relationship between L2 teacher cognition research and language assessment literacy research. I wanted to better understand what Uzbekistan EFL teachers perceive to be valuable in terms of knowledge and skills of assessment, what they do (with assessment), and why they do what they do. The dissertation more fully illuminated the conceptual clarity of L2 teachers' language assessment literacy with the analysis of social, cultural, historical, and political factors. Scarino (2013) iterated that it is necessary to consider not only the knowledge base in its most contemporary representation, but also to teachers' interpretive frameworks, "which are shaped through their particular situated personal experiences, knowledge, understanding and beliefs" (Scarino, 2013, p. 322). Based on the qualitative results, I have a new definition for language assessment literacy for classroom

language teachers that bridges the two fields. Language assessment literacy for language teachers is:

the capacity to be responsive to the fluidity of one's own cognitions on assessment/testing practices in order to create relevant and meaningful assessment experiences for learners that are situated within specific local sociocultural and sociopolitical contexts.

In other words, being language assessment literate is to have a connection and/or to build a connection between what teachers do, what they think about what they do, and how they assess students in a relevant, meaningful, and ethical way. Additionally, to be assessment literate a language teacher has to be able to make necessary changes in his or her assessment practices so that language learners may benefit (i.e., learn more effectively) from such changes.

9.6.3 Language Assessment Literacy

The dissertation contributes to the literature on language assessment literacy and supports Taylor's (2013) premise that different stakeholders in language assessment have varying needs. The results of the first research question provided some evidence for Kremmel and Harding's survey's validity. Additionally, the empirical results of the first research question, which includes the descriptive statistics and factor analysis, parallel Taylor's (2013) initial hypothesis. She explained that "a profile for classroom language teachers, however, may end up focusing strongly on the practical know-how needed for creating tests but have a much lighter focus on measurement theory or ethical principles; the latter may need to be touched upon only briefly at a surface level" (p. 409).

In addition to contributing to the literature on language assessment literacy, the results of the dissertation showed how there are implications for SLTE among language teachers (in

general) and Uzbekistan EFL teachers (in particular). These implications could move the field of language assessment literacy forward in three ways. First, our ways of understanding and defining the language assessment literacy construct can be better conceptualized. Second, we can understand how we might use language and discourse more effectively when engaging with a non-specialist audience in order to support assessment literacy development. Third, we can better observe how language assessment literacy grows and matures over time. I trust that one of the overarching contributions of the dissertation is that when research methods emerging from L2 teacher cognition research are joined with language assessment literacy research, a fuller understanding of ways of supporting the language assessment development of L2 teachers is revealed.

REFERENCES

- Acqandro, C., Jambon, B., Ellis, D., & Marquis, P. (1996). Language and translation issues. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trial* (2nd ed.), (pp. 575-585). Philadelphia, PA: Lippincott-Raven.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *The Standards for Competence in the Educational Assessment of Students*. Retrieved July 10, 2018, from <http://www.unl.edu/buros/article3.html>
- Andrade, H. (2013). Classroom assessment in the context of learning theory and research. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 17-34). New York: SAGE.
- Andrews, S. (1999). 'All these like little name things': a comparative study of language teachers' explicit knowledge of grammar and grammatical terminology. *Language Awareness*, 8 (3/4), 143-59.
- Bachman, L., & Dambök, B. (2018). *Language assessment for classroom teachers: Assessment for teachers*. Oxford: Oxford University Press.
- Bailey, K. M., & Curtis, A. (2015). *Learning about language assessment: Dilemmas, decisions, and directions* (2nd ed.), Boston, MA: Heinle Cengage Learning.
- Bailey, K. M. (2006). *Language teacher supervision: A case-based approach*. Cambridge: Cambridge University Press.
- Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, 48(1), 136-163. doi: 10.1002/tesq.99

- Basturkmen, H. (2012). Review of research into the correspondence between language teachers' stated beliefs and practices. *System*, 40, 282-295.
doi:10.1016/j.system.2012.05.001
- Basturkmen, H., Loewen, S., & Ellis, R. (2004). Teachers' stated beliefs about incidental focus on form and their classroom practices. *Applied Linguistics*, 25, 243-272.
- Beal, D. (2017, December). Uzbekistan: The tiled cities and vast deserts of the Central Asian country, once at the crossroads of civilization, are now thrillingly suspended between past and future. *Condé Nast Traveler*, pp. 56-69.
- Beaton, D. E., Bombardier, C., Guillemin, F., Ferraz, M. B. (2002). *Recommendations for the cross-cultural adaptation of health status measures* (2nd ed.). Rosemont, IL: American Academy of Orthopedic Surgeons. doi:
<https://www.ortho.umn.edu/sites/ortho.umn.edu/files/recommendations-cultural.pdf>
- Beauchamp, C., & Thomas, L. (2009). Understanding teacher identity: An overview of issues in the literature and implications for teacher education. *Cambridge Journal of Education*, 39(2), 175-189.
- Behr, D., & Shishido, K. (2016). The translation of measurement instruments for cross-cultural surveys. In C. Wolf, D. Joyce, T. Smith, & Y-Ch. Fu (Eds.), *The SAGE Handbook of Survey Methodology* (pp. 269-287). doi:
<http://dx.doi.org/10.4135/9781473957893.n19>
- Birenbaum, M. (2014). Conceptualizing assessment culture in school. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Assessment for learning improvement and accountability: The enabling power* (pp. 285-302). Netherlands: Springer.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through

- classroom assessment. *The Phi Delta Kappan*, 80(2), 139-144, 146-148.
<http://www.jstor.org/stable/20439383>
- Borg, S. (1999). Studying teacher cognition in second language grammar teaching. *System*, 27 (1), 19-31.
- Borg, S. (2003a.). Teacher cognition in language teaching: A review of research on what language teachers think, know, and do. *Language Teaching*, 36(2), 81-109. doi: 10.1017/S0261444803001903
- Borg, S. (2003b.). Teacher cognition in grammar teaching: A literature review. *Language Awareness*, 12(2), 96-108. doi: 10.1080/09658410308667069
- Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London, UK: Continuum. doi: 10.5040/9781474219983
<http://dx.doi.org/10.5040/9781474219983>
- Borg, S. (2009). Language teacher cognition. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide to second language teacher education* (pp. 163-171). Cambridge, UK: Cambridge University Press.
- Borg, S. (2012). Current approaches to language teacher cognition research: A methodological analysis. In R. Barnard & A. Burns (Eds.), *Researching language teacher cognition and practice: International case studies* (pp. 11-29). Bristol, UK: Multilingual Matters.
- Bourke, R., & Mentis, M. (2013). Self-assessment as a process for inclusion. *International Journal of Inclusive Education*, 17(8), 854-867.
- Breen, M. P., Hird, B., Milton, M., Oliver, R., & Thwaite, A. (2001). Making sense of language teaching: Teachers' principle and classroom practices. *Applied Linguistics*, 22, 470-501.

- Brown, J. D. (2001). *Using surveys in language programs*. Cambridge, UK: Cambridge University Press.
- Burns, A., & Knox, J. S. (2011). Classroom as complex adaptive systems: A relational model. *TESL-EJ*, 15, 1-25.
- Burns, A., Freeman, D., & Edwards, E. (2015). Theorizing and studying the language teaching mind: Mapping research on language teacher cognition. *The Modern Language Journal*, 99(3), 585-601. doi: 10.1111/modl.12245
- Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth, NH: Heinemann.
- Burri, M., Chen, H., & Baker, A. (2017). Joint emergence of teacher cognition and identity through learning to teach L2 pronunciation. *The Modern Language Journal*, 101(1), 128-142. doi: 10.1111/modl.12388
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
- Chiesa, D. L. (Recorder). (July 04, 2017). *Ambassador Spratlen's Opening Speech at 4th of July Celebration with Uzbekistan Government Officials*. [Apple iPhone Audio File].
- Clark, C., & Peterson, P. (1986). Teacher's thought processes. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching*. (3rd. ed.) (pp. 255-296). New York: New York: MacMillan.
- CEFR (Common European Framework of Reference for Languages). In: ManagementMania.com [online]. Wilmington (DE) 2011-2018, 08/22/2016 [cit. 07/18/2018]. Available at: <https://managementmania.com/en/common-european-framework-of-reference-for-languages-cefr>

- Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of Exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147–168.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O’Sullivan, & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge, UK: Cambridge University Press.
- Cross, R. (2010). Language teaching as sociocultural activity: Rethinking language teacher practice. *Modern Language Journal*, 94(3), 434-452. doi: 10.1111/j.1540-4781.2010.01058.x
- Cumming, A. (2003). Experienced ESL/EFL instructors; conceptualizations of their teaching: Curriculum options and implications. In B. Kroll (Ed.), *Exploring the dynamics of L2 writing* (pp. 71-93). Cambridge: Cambridge University Press.
- Cumming, A., & Berwick, R. (Eds.). (1996). *Validation in Language Testing*. Clevedon, England: Multilingual Matters, Ltd.
- de Bot, K. (2008). Introduction: Second language as a dynamic process. *Modern Language Journal*, 92(2), 166-178.
- Davies, A., Busick, K., Herbst, S. & Sherman, A. (2014). System leaders using assessment for learning as both the change and the change process: Developing theory from practice. *The Curriculum Journal*, 25(4), 567-592.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates’ learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419–438. doi: 10.1080/0969594X.2010.516643

- Denzin, N. K., & Lincoln, Y. S. (Eds.) (2005). *The Sage handbook of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Dillon, D. R. (2013). Grounded theory and qualitative research. *The Encyclopedia of Applied Linguistics*. Wiley Online Library. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781405198431>
- DiPardo, A., & Potter, C. (2003). Beyond cognition: A Vygotskian perspective on emotionality and teacher's professional lives. In A. Kozulin, B. Gindis, & S. M. Miller (Eds.), *Vygotsky's Educational Theory in Cultural Context* (pp. 317-345). Cambridge, UK: Cambridge University Press.
- Dixon-Woods, M., Cavers, D., Agarwal, S., Annanale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R., & Sutton, A. J. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups, *BMC Medical Research Methodology*, 6(35), [open access], <https://doi.org/10.1186/1471-2288-6-35>
- Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Dowsett, G. (1986). *Interaction in the semi-structured interview*. In M. Emery (ed.), *Qualitative research* (pp. 50-56). Canberra: Australian Association of Adult Education.
- Ellis, N., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics (introduction to special issue). *Applied Linguistics*, 27, 558-589.
- ESCCorg (2013, 08, 20). *Assessment literacy: The overlooked ingredient in educator effectiveness*. [Video file]. Retrieved from: https://www.youtube.com/watch?v=Wg_tz3GJy0U

- Falvey, P., & Cheng, L. (1995). A comparative study of teachers' beliefs about assessment principles and practices. *Language Testing Update*, 18, 38–39.
- Farrell, T. S. C. (1999). The reflective assignment: Unlocking preservice teachers' beliefs on grammar teaching. *RELC Journal*, 30, 1-17.
- Farrell, T. S. C. (2009). Critical reflection in a TESL course: Mapping conceptual change. *ELT Journal*, 63, 221-229.
- Feryok, A. (2008). An Armenian English language teacher's practical theory of communicative language teaching. *System*, 36(2), 227-240.
doi:10.1016/j.system.2007.09.004
- Feryok, A. (2010). Language teacher cognitions: Complex dynamic systems? *System*, 38, 272-279.
- Feryok, A., & Oranje, J. (2015). Adopting a cultural portfolio project in teaching German as a foreign language: Language teacher cognition as a dynamic system. *Modern Language Journal*, 99(3), 546-564.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Finch, A. (2010). Critical incidents and language learning: Sensitivity to initial conditions. *System*, 38(3), 422-431. doi: 10.1016/j.system.2010.05.004
- Freeman, D. (1989). Teacher training, development, and decision making. *TESOL Quarterly*, 23(1), 27-47. doi: 10.2307/3587506.
- Freeman, D. (1993). Renaming experience/reconstructing practice: Developing new understanding of teaching. *Teaching and Teacher Education*, 9, 485-497.
- Freeman, D. (2016). *Educating second language teachers: 'The same things done differently.'* Oxford: Oxford University Press.

- Freeman, D. & Johnson, K. (1998). Reconceptualizing the knowledge-base of language teacher education. *TESOL Quarterly*, 32(3), 397-417. doi: 10.2307/3588114
- Freeman, D., & Johnson, K. E. (2005). Towards linking teacher knowledge and student learning. In D. J. Tedick (Ed.), *Language teacher education: International perspectives on research and practice* (pp. 73-95). Mahwah, NJ: Lawrence Erlbaum.
- Freeman, D., & Ricards, J. C. (Eds.). (1996). *Teacher learning in language teaching*. Cambridge: Cambridge University Press.
- Freire, P. (1970). *Pedagogy of the oppressed*. New York: Herder and Herder.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113- 132. <http://dx.doi.org/10.1080/15434303.2011.642041>
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gibson, J. J. (1977). *The theory of affordances*. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Erlbaum.
- Gilje, T. M. (2014). Teacher cognition and the teaching of EFL reading in Norwegian upper primary classrooms. *Acta Didactica Norge*, 8(2), 1-17.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational measurement*. London: Falmer.
- Gleick, J. (1987). *Chaos: Making a new science*. New York: Viking Penguin.
- Golombek, P. R., & Doran, M. (2014). Unifying cognition, emotion, and activity in language teacher professional development. *Teaching and Teacher Education*, 39, 102-111. <http://dx.doi.org/10.1016/j.tate.2014.01.002>

- Grossman, P. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teacher College Press.
- Hair, J., Anderson, R. E., Tatham Black, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hakim, B. (2015). English language teachers' ideology of ELT assessment literacy. *International Journal of Education & Literacy Studies*, 4(3), 42-48.
doi:10.7575/aiac.ijels.v.3n.4p.42
- Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 487–504). Norwell, MA: Springer.
- Hantrais, L. (2009). *International comparative research: Theory, methods and practice*. Hampshire, UK: Palgrave MacMillan.
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment: Volume 12 of handbooks on applied linguistics* (pp. 413-428). Berlin, Germany: Walter de Gruyter GmbH & Co KG. (chapter doi):
<https://doi.org/10.1515/9781614513827-027>
- Harkness, J. A. (2008). Comparative survey research: Goals and challenges. In E. D. de Leuw, J. J., Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 56-77). New York/London: Lawrence Erlbaum Associates.
- Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European survey of language testing and a assessment needs. Part 1: General findings*. Gothenburg, Sweden: European Association for Language Testing and Assessment. Available from:

<http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf>

- Hatch, E., & Lazarson, A. (1991). *The research manual: Design and statistics for applied linguistics*. Heinle: Boston.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hill, K., & McNamara, T. (2011). Developing a comprehensive empirically based research framework for classroom based assessment. *Language Testing*, 29(3), 395-420. doi: 10.1177/0265532211428317
- Hill, M. F., Cowie, B., Gilmore, A., & Smith, L. F. (2010). Preparing assessment-capable teachers: What should pre-service teachers know and be able to do? *Assessment Matters*, 2, 44-64.
- Hiver, P. (2013). The interplay of possible language teacher selves in professional development choices. *Language Teaching Research*, 17(2), 210-227. doi: 10.1177/1362168813475944
- Ho, D. G. E. (November 05, 2012). *Focus groups*. In C. Chappelle's (Ed.), *The Encyclopedia of Applied Linguistics*.
<https://doi-org.ezproxy.gsu.edu/10.1002/9781405198431.wbeal0418>
- Huta, A., Hirvala, T., Banerjee, J. (2005). *European survey of language testing and a assessment needs. Part 2: Regional findings*. Gothenburg, Sweden: European Association for Language Testing and Assessment. Available from:
http://users.jyu.fi/~huhta/ENLT2/First_page.html
- In'nami, Y. & Koizumi, R. (2011). Structural equation modeling in language testing and learning research: A review. *Language Assessment Quarterly*, 8(3), 250-276.

- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. doi: 10.1177/0265532208090158
- Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2923–2931). Oxford: Blackwell.
- Islam Karimov Sayings [Museum exhibit label]. Islam Karimov Museum, Tashkent, Uzbekistan.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584. doi: 10.1177/0265532213480334
- Kaplowitz, M. D. (2000). Statistical analysis of sensitive topics of group and individual interviews. *Quality and Quantity*, 34(4), 419-431.
- Johnson, K. E. (2009). *Second language teacher education: A sociocultural perspective*. New York/London: Routledge.
- Johnson, K. E., & Golombek, P. R. (2016). *Mindful L2 teacher education: A sociocultural perspective on cultivating teachers' professional development*. New York: Routledge.
- Kaiser, G., & Willander, T. (2005). Development of mathematical literacy: Results of an empirical study. *Teaching mathematics and its applications*, 24(2/3), 48-60.
- Kanno, Y., & Stuart, C. (2011). Learning to become a second language teacher: Identities-in-practice. *Modern Language Journal*, 95, 236-252.
- Kaplowitz, M. D. (2000). Statistical analysis of sensitive topics of group and individual interviews. *Quality and Quantity*, 34(4), 419-431.
- Karavas-Doukas, E. (1996). Using attitude scales to investigate teachers' attitudes to the

- communicative approach. *ELT Journal*, 50(3), 187-198. doi: 10.1093/elt/50.3.187
- Kreuger, R. A. (1994). *Focus groups interviews: A practical guide for applied research* (2nd ed.). Thousand Oaks, Ca.: Sage Publications.
- Kaiser, H. 1974. An index of factor simplicity. *Psychometrika*, 39, 31–36.
- Kim, J. I. (2015). Developing and validating a survey of Korean early childhood English teachers' knowledge. *English Language Teaching*, 8(12), 170-181.
- Kiss, T. (2012). The complexity of teacher learning: Reflection as a complex dynamic system. *Journal of Interdisciplinary Research in Education*, 2(1), 17-35.
- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255–276. doi: 10.1080/10476210.2011.593164
- Kramer, A. (2017, October). Once closed and repressive, Uzbekistan is opening up. *New York Times Magazine*, retrieved from, <https://www.nytimes.com/2017/10/25/world/asia/uzbekistan-politics.html>.
- Krause, K.-L., Bochner, S., & Duchesne, S. (2006). *Educational psychology for learning and teaching*. (2nd ed.) South Melbourne, Victoria, Australia: Thomson.
- Kremmel, B., & Harding, L. (forthcoming). *Towards a comprehensive, empirical model of language assessment literacy*.
- Kremmel, B., & Harding, L. (2016). *Pilot results from a validation study*. Language Assessment Literacy Symposium Poster Presentation.
- Kreuger, R. A. (1994). *Focus groups interviews: A practical guide for applied research* (2nd ed.). Thousand Oaks, Ca.: Sage Publications.
- Kubanyiova, M. (2012). *Teacher development in action: Understanding language teachers' conceptual change*. Basingstoke, UK: Palgrave MacMillan.

- Kubanyiova, M., & Feryok, A. (2015). Language teacher cognition in applied linguistics research: Revisiting the territory, redrawing the boundaries, reclaiming the relevance. *The Modern Language Journal*, 99(3), 435-449. doi: 10.1111/modl.12239
- Kumaravadivelu, B. (1994). The postmethod condition: (E)merging strategies for second/foreign language teaching. *TESOL Quarterly*, 28(1), 27-48. doi:10.2307/3587197
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197. 10.1177/0265532214554321
- Lange, D. C. (1990). A blueprint for a teacher development program. In Jack C. Richards and David Nunan (Eds.), *Second language teacher education* (pp. 245-268). Cambridge: Cambridge University Press.
- Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18, 141–165.
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. New York/Oxford: Oxford University Press.
- Lee, J. J., Murphy, J., & Baker, A. (2015). “Teachers are not empty vessels”: A reception study of Freeman and Johnson’s (1998) Reconceptualization of the knowledge base of second language teacher education. *TESL Canada Journal/Revue TESL Du Canada*, 33(1), 1-21. doi: <http://dx.doi.org/10.18806/tesl.v33i1.1224>
- Lett, J. (1990). Emics and etics: Notes on the epistemology of anthropology. In Thomas N. Headland, K. L. Pike, and M. Harris (Eds.), *Emics and etics: The insider/outsider debate* (pp. 127-142). Newbury Park, CA: Sage.
- Lietz, P. (2008). *Questionnaire design in attitude and opinion research: Current state of an art*. Bremen, Germany: Working Papers FOR655.

- Likert, R. (1932). A technique for the measurement of attitude scales. *Archives of Psychology*, 22(140), 1-55.
- Lincoln, M. H., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Lipa, S. E., & Harlin, R. (1990). *Assessment: Insights into teachers' beliefs and practices about process writing*. Paper presented at the Annual Meeting of Eastern Educational Research Association, Clearwater, FL. Available from the Educational Resources Information Center (ERIC) database.
- Liu, Y., & Fisher, L. (2006). The development patterns of modern foreign language student teachers' conceptions of self and their explanations about change: Three cases. *Teacher Development*, 10(3), 343-360. doi: 10.1080/13664530600922203
- Loewen, S., & Gonulal, T. (2015). Factor analysis. In S. Loewen, & L. Plonsky, L. (Eds.), *An A-Z of applied linguistics research methods* (p. 182-212). New York, NY: Palgrave.
- Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- MacDonald, M., Badger, R., & White, G. (2001). Changing values: What use are theories of language learning and teaching? *Teaching and Teacher Education*, 17, 949-963.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Madriz, E. (2000). Focus groups in feminist research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed.) (pp. 835-850). Thousand Oaks, Ca.: Sage Publications.
- MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education*, 20(5), 523-535.

- Madriz, E. (2000). Focus groups in feminist research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed.) (pp. 835-850). Thousand Oaks, Ca.: Sage Publications.
- Mahboob, A., Paltridge, B., Phakiti, A., Wagner, E., Starfield, S., Burns, A., & Jones, R. H. (2016). TESOL Quarterly research guidelines, *TESOL Quarterly*, 50(1), 41-65.
- Malone, M. E. (2008). Training in language assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education*, Vol. 7: Language testing and assessment (2nd ed.), (pp. 225-233). New York, NY: Springer Science and Business.
- Mattheoudakis, M. (2007). Tracking changes in pre-service EFL teacher beliefs in Greece: A longitudinal study. *Teaching and Teacher Education*, 23, 1272-1288. doi: 10.1016/j.tate.2006.06.001
- McMillan, J. H., & Schumacher, S. (2001). *Research in education: A conceptual introduction*. (5th ed.). New York: Addison Wesley Longman.
- McMillan, J. H. (2013). *SAGE handbook of research on classroom assessment*. New York, NY: Sage.
- McNeill, D. (Ed.) (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Mertler, C. A. (2003). *Classroom assessment: A practical guide for educators*. Los Angeles, CA: Pyrczak.
- Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montréal, Quebec, Canada April 11–15, 2005.
- Mohler, P., Dorer, B., de Jong, J., & Hu, M. (2016). Translation overview. *Guidelines for*

- Best Practice in Cross-Cultural Surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved 01,12, 2017, from <http://www.ccsr.isr.umich.edu/>
- Mullock, B. (2006). The pedagogical knowledge base of four TESOL teachers. *The Modern Language Journal*, 90(1), 48-65. doi: 10.1111/j.1540-4781.2006.00384.x
- Nunan, D., & Bailey, K. M. (2009). *Exploring second language classroom research: A comprehensive guide*. Boston, MA: Heinle, Cengage Learning.
- Numrich, C. (1996). On becoming a language teacher: Insights from diary studies. *TESOL Quarterly*, 30, 131-153. doi: 10.2307/3587610
- O'Loughlin, K. (2006). Learning about second language assessment: Insights from a post-graduate student online forum. *University of Sydney Papers In TESOL*, 1(1), 75-85. Retrieved from: http://faculty.edfac.usyd.edu.au/projects/usp_in_tesol/pdf/volume01/article04.pdf
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 307-332.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Pennell, B.-E., Cibelli-Hibben, K. L., Lyberg, L., Mohler, P. P., & Worku, G. (2017). A Total Survey Error Perspective on Surveys in Multinational, Multiregional, and Multicultural Contexts. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter,

- L. Lyberg, C. Tucker, & B. West (Eds.), *Total Survey Error in Practice*. New York, NY: John Wiley & Sons.
- Petric, B., & Czarl, B. (2003). Validating a writing strategy questionnaire. *System, 31*, 187-215. doi:10.1016/S0346-251X(03)00020-4
- Pike, K. L. (1954). *Language in relation to a unified theory of the structure of human behavior. Part I*. (Preliminary ed.). Glendale, CA: The Summer Institute of Linguistics.
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(3), 381-402. doi: 10.1177/0265532213480337
- Plake, B. S., & Impara, J. C. (1996). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achieving and adjustment* (pp. 53-68). New York, NY: Academic Press.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10-12, 39.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or Fundamental? *Theory Into Practice, 48*(1), 4-11. doi: 10.1080/00405840802577536
- Popham, W. J. (2011, February). *Formative assessment a process, not a test*. EducationWeek. Retrieved August 10, 2018, from http://www.edweek.org/ew/articles/2011/02/23/21popham_h30.html?t
- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*, 115e131.
- Reid, J. (1990). The dirty laundry of ESL survey research. *TESOL Quarterly, 24*(2), 323-338. doi: 10.2307/3586913

- Richards, J. C. (1996). Teachers' maxims in language teaching. *TESOL Quarterly*, 30(2), 281-296. doi: 10.2307/3588144
- Richards, J. C., Ho, B., & Giblin, K. (1996). Learning how to teach in the RSA cert. In D. Freeman, & J. C. Richards (Eds.), *Teacher Learning in Language Teaching* (pp. 242-259). Cambridge, MA: Cambridge University Press.
- Richards, J. C. & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. Cambridge: Cambridge University Press.
- Richards, J. C., & Nunan, D. (1990). *Second language teacher education*. Cambridge, UK: Cambridge University Press.
- Rietveld, T., & Van Hout, R. (1993). *Statistical techniques for the study of language and language behavior*. New York: Mouton de Gruyter.
- Saris, W., van der Veld, W., & Gallhofer, I. (2004). Development and improvement of questionnaires using predictions of reliability and validity. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires*. New York, NY: John Wiley and Sons.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309-327. doi: 10.1177/0265532213480128
- Schweikard, D. P., & Schmid, H. B. (2013). Collective intentionality. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved on January 10, 2017 from <https://plato.stanford.edu/>.
- Searle, J. R. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. E.

- Pollack (Eds.), *Intentions in communication* (pp. 401–415). Cambridge, MA: The MIT Press.
- Seliger, H., & Shohamy, E. (1989). *Second language research methods*. Oxford University Press: Oxford.
- Shavelson, R. (1973). What is the basic teaching skill? *Journal of Teacher Education*, 22, 144-151. doi: 10.177/002248717302400213
- Smith, L. T. (2013). *Decolonizing research methodologies: Shared communities and practices*. Colorado State University, Denver: Colorado.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
<http://www.jstor.org/stable/20404455>
- Stiggins, R. J. (2002). The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 785-765. <http://www.jstor.org/stable/20440249>
- Svalberg, A. M-L. (2015). Understanding the complex process in developing student teachers' knowledge about grammar. *Modern Language Journal*, 99(3), 529-545.
- Swain, M. (1984). Teaching and testing communicatively. *TESL Talk* 15, 1 and 2, 7-18.
- Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36. doi:10.1017/S0267190509090035
- Taylor, L. (2013). Communicating the theory, practice, and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. doi: 10.1177/0265532213480338
- Tsui, A. B. M. (1996). Learning how to teach ESL writing. In D. Freeman & J. C. Richards

- (Eds.), *Teacher learning in language teaching* (pp. 97-119). Cambridge: Cambridge University Press.
- Tsui, A. B. M. (2003). *Understanding expertise in teaching: Case studies of second language teachers*. Cambridge, UK: Cambridge University Press.
- Urmston, A. (2003). Learning to teach English in Hong Kong: The opinions of teachers in training. *Language and Education*, 17, 112-137. doi: 10.1080/09500780308666843
- van der Veer, R., & Valsiner, J. (1994). *The Vygotsky reader*. Cambridge: Blackwell.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770. doi: 10.2307/20466661
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1994). The problem of the environment. In R. van der Veer, & J. Valsiner (Eds.), *The Vygotsky reader* (pp. 338-354). Oxford: Blackwell Publishers.
- Wagenmakers, E. J. (2018). *JASP v.0.8.3.1* [Computer software]. Amsterdam, The Netherlands: University of Amsterdam.
- Walberg, H. (1972). Decision and perception: New constructs in research on teaching effects. *Cambridge Journal of Education*, 7, 12-20.
- Warford, M. K., & Reeves, J. (2003). Falling into it: Novice TESOL teacher thinking. *Teachers and Teaching*, 9(1), 47-66. doi: 10.1080/1354060032000049904
- Warwick, D. P., & Lininger, C. A. (1975). *The sample survey: Theory and practice*. Texas: McGraw Hill.
- Wolf, C., Joy, D. Smith, T., & Fu, Y. (Eds.). (2016). *The Sage handbook of survey*

- methodology*. Thousand Oaks, CA: Sage Publications.
- Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making and classroom practice*. Cambridge, UK: Cambridge University Press.
- Wright, T. (2010). Second language teacher education: Review of recent research on practice. *Language Teaching*, 43(3), 259-296. doi:10.1017/S0261444810000030
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.
<http://dx.doi.org/10.1016/j.tate.2016.05.010>
- Xu, Y & Liu, Y (2009) Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 492-513.
- Yuan, R. & Lee, I. (2014). Pre-service teachers' changing beliefs in the teaching practicum. Three cases in an EFL context. *System*, 44, 1-12.
- Zolfaghari, S., & Ashraf, H. (2015). The relationship between EFL teachers' assessment literacy; their teaching experience; and their age: A case of Iranian EFL teachers. *Theory and Practice in Language Studies*, 5(12), 2550-2556.
<http://dx.doi.org/10.17507/tpls.0512.16>

APPENDICES

Appendix A: Uzbekistan Presidential Decree, No. 24

DECREE

OF PRESIDENT OF THE REPUBLIC OF UZBEKISTAN

ABOUT MEASURES FOR FURTHER IMPROVEMENT OF THE SYSTEM OF RETRAINING AND ADVANCED TRAINING OF MANAGERIAL AND PEDAGOGICAL PERSONNEL IN HIGHER EDUCATIONAL INSTITUTIONS

(Collection of laws of the Republic of Uzbekistan, 2015, № 24, p. 312)

With a view to dramatically improve the quality of training the highly qualified specialists on the basis of continuous development of professional level and qualification of the faculty in higher educational institutions, and introduction of improved system of their regular retraining in accordance with modern requirements:

1. The following shall be considered as the most important directions of further improvement the system of retraining and advanced training of managerial and pedagogical personnel in higher educational institutions:

improving on a regular basis the pedagogical and professional level of the faculty in higher educational institutions, in-depth study by them of legislation norms, the recent advances in theory, scientific and applied researches, technological progress and innovations on teaching discipline, as well as modern methods of educational process organization;

radical renewal of qualification requirements, curricula, programmes and methods of retraining and advanced training of the faculty in higher educational institutions with due account of widespread introduction of modern high-performance educational and innovative technologies, and advanced international experience;

acquisition by professors in higher educational institutions and the active introduction in educational process of modern innovative pedagogical, information and communication technologies and the global network of Internet, multimedia systems and distance learning techniques;

raising the level of practical knowledge of foreign languages by the faculty in higher educational institutions and the wide use of this knowledge for steady growth of their professional skills in teaching and research activities.

2. Proposal shall be accepted made by the Ministry of Higher and Secondary Specialized Education, Ministry of Finance of the Republic of Uzbekistan, the Higher Attestation Commission at the Cabinet of Ministers, other ministries and agencies having under their supervision the higher educational institutions, to define 15 leading higher educational institutions in the country as per [Annex 1](#) as the Base higher educational institutions for organization of retraining and advanced training of managerial and pedagogical personnel in higher educational institutions in areas of retraining, which have on these directions the Academic Councils for Doctoral Theses, the highly qualified highly qualified scientific and pedagogical potential, modern, equipped with necessary methodical, training and laboratory, information and communication means, and material-technical base.

The main tasks and functions of the Base higher educational institutions shall be set as following:

organization of constantly operating courses for retraining and advanced training of managerial and teaching personnel in higher educational institutions on relevant areas of retraining;

creation of necessary educational, methodical and material-technical base to conduct training courses at qualitatively high organizational and professional level;

wide attraction of leading professors from higher educational institution and members of the Scientific Council for Doctoral Theses, as well as highly qualified faculty from other universities in the country, and foreign experts, educational professionals and practitioners on contractual basis to conduct classes at courses for retraining and advanced training;

forming the necessary information and reference base in the areas of retraining, development and introduction in retraining process of modern innovative educational, multimedia and information and communication technologies;

organization of pedagogical practice for trainees of retraining and advanced training courses where they shall conduct open lectures and practical classes followed by with discussion and critical analysis;

taking measures to provide nonresident trainees with accommodation in dormitories of higher educational institutions for the period of retraining and advanced training.

The rectors of higher educational institutions defined as the Base higher educational institutions for retraining and advanced training of managerial and pedagogical personnel of higher educational institutions shall be imposed the personal responsibility for organization of the retraining on high quality level.

3. The procedure shall be established in accordance with which:

managerial and pedagogical personnel of higher educational institutions in the country, regardless of their departmental affiliation, are required on regular basis, at least once in three years, to undergo retraining and advanced training in relevant areas of retraining on permanent courses at Base higher educational institutions as per [Annex 1](#);

retraining and advanced training courses shall be arranged for two months on off-the-job conditions under the special programs for 288 hours, developed on the basis of updated Model structure of retraining curriculum for managerial and pedagogical personnel in higher educational institutions, as per [Annex 2](#);

upon completion of training, the trainees are subject to certification held by Certification commissions established in each Base higher educational institution and headed its rector or the President of Academic Council for Doctoral Theses composed of at least 7 people from among the leading scientists and specialists in their respective fields, members of the Academic Council, the major specialists in the field of educational and methodical work;

trainees who successfully went through the certification, shall be issued the Qualification certificate as per [Annex 3](#), which is the nationally recognized document of strict reporting;

managerial and pedagogical personnel of higher educational institutions who failed to pass the certification after the retraining and advanced training courses must within the set deadlines go through the repeated retraining and certification;

if the managerial and pedagogical personnel of higher educational institutions fail again to pass the re-certification they will lose the right to perform teaching activity in higher educational institutions and at the main place of their employment their labor contract shall be terminated;

pedagogical personnel successfully passed the certification after undergoing the training and advanced training courses have the priority right at competition for vacant teaching positions in the relevant field;

trainees who undergo the retraining and advanced training courses in Base higher educational institutions shall remain employed for the whole period of training on their position at the main place of employment and the average salary.

Organization the activity of constantly operating courses for retraining and advanced training of managerial and pedagogical personnel of higher educational institutions shall be started from September 1, 2015.

4. Ministry of Higher and Secondary Specialized Education shall before 1 August 2015 ensure the development and approval in established procedure of updated qualification requirements, curricula and programs of retraining and advanced training of pedagogical personnel of higher education institutions, with due account of Model structure of curriculum for retraining course of managerial and pedagogical personnel of higher educational institutions, as well as modern requirements.

5. In order to ensure clear and effective coordination of works on organization of processes of retraining and certification of managerial and pedagogical personnel of higher educational institutions in the country, there shall be established the permanent Inter-agency Commission under the Cabinet of Ministers of the Republic of Uzbekistan with imposing on it of the following tasks:

organization of all work and ensuring proper control over implementation of the provisions on further improvement of the system of retraining and advanced training of managerial and pedagogical personnel of higher educational institutions envisaged in this Decree, paying the special attention to increasing of interest among faculty members in continual growth of their professional skills in accordance with modern requirements;

coordination the activity of Base higher educational institutions on organization of retraining and advanced training of managerial and pedagogical personnel of higher educational institutions, providing assistance in further development and strengthening of their educational-methodical and material-technical base;

organization of work for continuous improvement with due account of modern requirements, curricula and programs of retraining and advanced training of pedagogical personnel of higher educational institutions in the areas of retraining;

for sustainable regulation of the process of retraining of pedagogical personnel, profession-oriented chairs of higher educational institutions, the faculty of which are undergoing the retraining and advanced training in the relevant fields of retraining, shall be attached to the specific Base higher educational institutions, meaning the establishment of regular relations between the Base higher educational institutions and the relevant professional chairs in higher educational institutions of the country;

forming and approval of the composition of Certification Commissions in Base higher educational institutions for organization of retraining and advanced training of management and pedagogical personnel of higher education institutions, with mandatory annual hearing of their reports and, if necessary, changing their composition.

6. To agree with proposal of the Ministry of Finance and the Ministry of Higher and Secondary Specialized Education of the Republic of Uzbekistan on increasing for 30% the salary of Presidents of Certification Commissions at Base higher educational institutions on organization of retraining and advanced training of managerial and pedagogical personnel.

At this it shall be set up that Presidents of Certification Commissions bear the personal responsibility for the quality level of certification of managerial and pedagogical personnel who underwent the retraining and advanced training, and the impartiality, transparency and adherence to principles at making decisions.

7. The Cabinet of Ministers of the Republic of Uzbekistan shall within one month adopt the resolution in pursuance of this Decree envisage in it the following:

approval of Provisions on retraining and advanced training courses for managerial and pedagogical personnel of higher educational institutions;

approval of composition of Inter-agency Commission for coordination of works on organization of processes of retraining and certification of managerial and pedagogical personnel of higher educational institutions;

measures for strengthening of teaching and methodical, and material and technical base in Base higher educational institutions, for equipping them with modern information and communication technologies and technical means of training.

8. Control over execution of this Decree shall be imposed on the Prime Minister of the Republic of Uzbekistan Sh.M.Mirziyoyev.

President of the Republic of Uzbekistan I. KARIMOV

Tashkent city,

12 June 2015,

№ UP-4732

Appendix B: GSU Internal Review Board Approval

INSTITUTIONAL REVIEW BOARD

Mail: P.O. Box 3999
Atlanta, Georgia 30302-3999
Phone: 404/413-3500
Fax: 404/413-3504

In Person: Dahlberg Hall
30 Courtland St, Suite 217



June 07, 2017

Principal Investigator: John Murphy

Key Personnel: Belcher, Diane; Chiesa, David L, Applied Linguistics and ESL; Coleman, Susan; Cushing, Sara; Murphy, John

Study Department: Applied Linguistics & ESL, Georgia State University

Study Title: Discerning assessment literacy as a situated social practice in Uzbekistan: Ingredients for a national in-service teacher education program

Review Type: Expedited, 6, 7

IRB Number: H17592

Reference Number: 344247

Approval Date:

06/07/2017

Expiration Date:

06/06/2018

The Georgia State University Institutional Review Board (IRB) reviewed and approved the above referenced study in accordance with 45 CFR 46.111. The IRB has reviewed and approved the study and any informed consent forms, recruitment materials, and other research materials that are marked as approved in the application. The approval period is listed above. Research that has been approved by the IRB may be subject to further appropriate review and approval or disapproval by officials of the Institution.

Appendix C: Adapted Survey in English

Part I

How **knowledgeable** do language teachers *need to be* about each aspect of language assessment below? Please respond according to the following scale:

0 = not knowledgeable at all

1 = slightly knowledgeable

2 = moderately knowledgeable

3 = very knowledgeable

4 = extremely knowledgeable

Item #		0	1	2	3	4
1	how to use assessments to inform learning or teaching goals					
2	how to use assessments to evaluate progress in language learning					
3	how to use assessments to evaluate achievement in language learning					
4	how to use assessments to diagnose learners' strengths and weaknesses					
5	how to use assessments to motivate student learning					
6	how to use self- and peer-assessment					
7	how to interpret assessment results appropriately					
8	how to interpret measurement error					
9	how to interpret what a particular score says about an individual's language ability					
<u>Any Clarification Needed?</u>						
Item #		0	1	2	3	4

10	how to determine if a local assessment aligns with a local education system					
11	how to determine if the content of a language assessment is culturally appropriate					
12	how to determine if the results of the assessment are relevant to the local context					
13	how to communicate assessment results and decisions to teachers					
14	how to communicate assessment results and decisions to students or parents					
15	how to determine whether a language assessment is useful for a particular purpose					
16	how to recognize when an assessment is being used inappropriately					
17	how to prepare learners to take language assessments					
<u>Any Clarification Needed?</u>						
Item		0	1	2	3	4
18	how to find information to help in interpreting results					
19	how to give useful feedback on the basis of assessment					
20	how assessments can influence the design of a language course or curriculum					
21	how assessments can influence teaching and learning materials					
22	how assessments can influence teaching and learning in the classroom					
23	how language skills develop					
24	how foreign/second languages are learned					
25	how language is used in society					
26	how social values can influence language assessment design and use					
<u>Any Clarification Needed?</u>						

Item		0	1	2	3	4
27	how pass-fail marks or cut-scores are set					
28	the concept of reliability					
29	the concept of validity					
30	the structure of language					
31	the advantages and disadvantages of standardized testing					
32	the philosophy behind the design of a relevant language assessment					
33	the impact language assessments can have on society					
34	the relevant legal regulations for assessment in the local area					
35	the assessment traditions in a local context					
<u>Any Clarification Needed?</u>						
Item #		0	1	2	3	4
36	the specialist terminology related to language assessment					
37	different language proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])					
38	different stages of language proficiency					
39	different types of purposes for language assessment purposes (e.g., proficiency, achievement, diagnostic)					
40	different forms of alternative assessments (e.g., portfolio assessment)					
41	one's own beliefs/attitudes towards language assessment.					
42	how one's own beliefs/attitudes might influence one's assessment practices					

43	how one's own beliefs/attitudes may conflict with those of other groups involved in assessment					
44	how one's own knowledge of language assessment might be further developed					
<u>Any Clarification Needed?</u>						

Part II

How skilled do language teachers need to be about each aspect of language assessment below? Please respond according to the following scale:

0 = not skilled at all

1 = slightly skilled

2 = moderately skilled

3 = very skilled

4 = extremely skilled

<u>Item #</u>		<u>0</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
45	using statistics to analyze the difficulty of individual items (questions)					
46	using statistics to analyze overall scores on a particular assessment					
47	using statistics to analyze the quality of individual items (questions)/tasks					
48	using techniques other than statistics (e.g., questionnaires, interviews, analysis of language) to get information about the quality of language assessment					
49	using rating scales to score speaking or writing performances					
50	using specifications to develop items (questions) and tasks					

51	scoring closed-response questions (e.g., Multiple Choice Questions)					
52	scoring open-ended questions (e.g., short answer questions)					
53	developing portfolio-based assessments					
54	developing specifications (overall plans) for language assessments					
55	selecting appropriate rating scales (rubrics)					
<u>Any Clarification Needed?</u>						
Item		0	1	2	3	4
56	Selecting appropriate items or tasks for a particular assessment purpose.					
57	Training others to use rating scales (rubrics) appropriately					
58	Training others to write good quality items (questions) or tasks for language assessment.					
59	Writing good quality items (questions) or tasks for language assessments.					
60	Aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])					
61	Determining pass-fail marks or cut-scores					
62	Identifying assessment bias.					
63	Designing scoring keys and rating scales (rubrics) for assessment tasks.					
64	Making decisions about what aspects of language to assess.					
65	Piloting/trying-out assessments before their administration					
66	Selecting appropriate ready-made assessments.					

Any Clarification Needed?

PART III – Biographical Information

GENDER: _____

AGE: _____

YEARS OF EXPERIENCE AS AN ENGLISH LANGUAGE TEACHER AT

UNIVERSITY: _____

ANY OTHER SUGGESTIONS FOR THIS SURVEY?

Appendix D: Survey in Uzbek and Russian

O'zbekistondagi ijtimoiy amaliyot paytida farq qiladigan savodxonlikni baholash: Milliy xizmatdagi o'qituvchi ta'lim dasturi uchun tarkibiy qismlar.

1-Qism

Quyida berilgan tilni baholashni har bir nuqtai-nazari to'g'risida til o'qituvchilari qanchalik bilimga ega bo'lishlari kerak? Iltimos quyidagi shkalaga mos holda javob bering:

- 0 = umuman bilimga ega emas
- 1 = ozgina bilimli
- 2 = o'rtacha bilimli
- 3 = chuqur bilimli
- 4 = o'ta darajada chuqur bilimli

Band		0	1	2	3	4
<u>1</u>	O'qitish va o'rganish maqsadlari haqida ma'lumot berishda baholashni qanday qo'llash					

<u>2</u>	Til o'rganishdagi rivojlanishga baho berish uchun baholashni qanday qo'llash					
<u>3</u>	Til o'rganishdagi yutuqlarga baho berish uchun baholashni qanday qo'llash					
<u>4</u>	O'rgatuvchilarni kuchlilik va zaiflik tomonlarini tashxislash uchun baholashni qanday qo'llash					
<u>5</u>	Talabani o'rganishga undash uchun baholashni qanday qo'llash					
<u>6</u>	O'zini va tengdoshlarini qanday baholash					
<u>7</u>	Baholash natijalarini mos holda qanday talqin qilish					
<u>8</u>	O'lchov xatosini qanday tahlil qilish					
<u>9</u>	Qisman ballni shaxsni tilga qobiliyati to'g'risida nima deyishini qanday tahlil qilish					

Biror oydinlashtirish kerakmi?

Band						
<u>10</u>	Mahalliy baholash mahalliy ta'lim tizimi bilan yonma-yon ekanligini qanday aniqlash					
<u>11</u>	Tilni baholashni mohiyati madaiyatga mos kelish yoki kelmasligini qanday aniqlash					
<u>12</u>	Baholash natijalari mahalliy vaziyatga aloqador ekanligini qanday aniqlash					

<u>13</u>	Qay tarzda baholash natijalari va xulosalarni o'qituvchilarga etkazish					
<u>14</u>	Qay tarzda baholash natijalari va xulosalarni talabalarga yoki ota-onalarga erkazish					
<u>15</u>	Tilni baholash ma'lum bir maqsadlar uchun foyda berish yoki bermasligini qanday aniqlash					
<u>16</u>	Baholash nomuvofiq ishlatilayotgan vaqtda qanday e'tirof etish					
<u>17</u>	O'rganuvchilarni tilni baholash sinovini topshirishga qanday tayyorlash					

Oydinlashtirish kerakmi ?

Band		0	1	2	3	4
<u>18</u>	Talqin qilish natijalariga yordam beruvchi ma'lumotlarni qanday topish					
<u>19</u>	Baholash asosida foydali fikr xulosalarni qanday berish					
<u>20</u>	Til kursi yoki ta'lim dasturi loyihasiga baholashning qanday ta'sir ko'rsatishi					
<u>21</u>	O'qitish va o'rganish materiallariga baholashning qanday ta'sir ko'rsatishi					
<u>22</u>	Sinf xonasida o'qitish va o'rganishga baholashni qanday ta'sir ko'rsatishi					
<u>23</u>	Tilga oid bilimlar qanday rivojlanadi					

<u>24</u>	chet/ikkinchi til qanday o'rganiladi					
<u>25</u>	Jamiyatda til qanday ishlatiladi					
<u>26</u>	Ijtimoiy qiymat tilni baholash loyahasiga va ishlatilishiga qanday ta'sir etadi					

Biror ovdinlashtirish kerakmi?

Band		0	1	2	3	4
<u>27</u>	O'tish yoki o'tolmaslik, yoki ballni olib tashlash qanday belgilanadi					
<u>28</u>	Ishonchlilik tushunchasi					
<u>29</u>	Yaroqlilik tushunchasi					
<u>30</u>	Til strukturasi (tuzilishi)					
<u>31</u>	Standartlashgan test sinovlarini ustunlik va zaif tomonlari					
<u>32</u>	Muvofiq tilni baholash loyihasi ortida falsafa yotadi					
<u>33</u>	Tilni baholash jamiyatga ta'sir ko'rsatishi mumkin					
<u>34</u>	Baholash uchun mahalliy tumanlardagi muvofiq qonuniy boshqaruvlar					

<u>35</u>	Mahalliy vaziyatdagi baholash an'analari						
------------------	--	--	--	--	--	--	--

<u>Biror ovdinlashtirish kerakmi?</u>
--

Band		0	1	2	3	4
<u>36</u>	Mutaxassis terminalogiasi tilini baholashga aloqador					
<u>37</u>	Turli xil tilga ixtisoslashgan tuzilmalar (masalan, the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])					
<u>38</u>	Tilga ustaligining har- xil bosqichilari					
<u>39</u>	Tilini baholash maqsadlari uchun maqsalranning har xil turlari (masalan: tajribalilik, erishish, tashxisiy)					
<u>40</u>	Muqobil baholashlarini har xil shakllari (masalan: umumiy-shaxsan baholash)					
<u>41</u>	Kimningdir tillini baholashga qaratilgan xususiy ishonchlari/munosabatlari					
<u>42</u>	Kimningdir xususiy ishonchlari/munosabatlari boshqa birovning baholash tajribasiga qanday ta'sir ko'rsatishi mumkinligi					
<u>43</u>	Kimningir xususiy ishonchlari/munosabatlari baholashda qatnashgan boshqa guruhlar bilan qanday ziddiyat keltirib chiqarishi mumkinligi					
<u>44</u>	Birovning tilini baholashdagi xususiy bilimlari keyinchalik rivojlantirilishi mumkinligi					

Biror oydinlashtirish kerakmi?

2-qism

Siz tanlagan guruh/kasb odamlari quyidagi tilini baholash aspektlari to'g'risida qay darajada tajribaga ega bo'lishlari kerak? Iltimos, quyidagi shkalaga mos holda javvob bering:

- 0 = tajriba kerak emas
 1 = kam tajribali
 2 = o'rtacha tajribali
 3 = juda tajribali
 4 = o'ta darajada tajribali

<u>Item #</u>		0	1	2	3	4
<u>45</u>	Yakka tartibdagi savollar mushkulligini tahlil qilish uchun statistikani qo'llash					
<u>46</u>	Qisman baholashdagi umumiy ballarni tahlili uchun statistikani qo'llash					
<u>47</u>	Yakka tartibdagi savollar/vasifalar sifatini tahlil qilishda statistikani qo'llash					

<u>48</u>	Tilni baholash sifati to'g'risida ma'lumot olish uchun statistikadan boshqa usullarni qo'llash (masalan: so'rov, suhbat o'tgazish, til tahlili)					
<u>49</u>	Gaplashish yoki yozish ijrosini baholashda reyting shkalasini qo'llash					
<u>50</u>	Savollar va vazifalarni rivojlantirishda o'ziga xos xususiyatlarni qo'llash					
<u>51</u>	Yopiq-javobli savollarni baholash (masalan: bir nechta javobli savollar)					
<u>52</u>	Ochiq- tugallanmagan savollarni baholash (masalan: qisqa javob beriladigan savollar)					
<u>53</u>	Shaxsga asosli baholashni rivojlantirish					
<u>54</u>	Tilni baholashda o'ziga xos xususiyatlarni (umumiy rejalar) rivojlantirish					
<u>55</u>	Muvofiq reyting shkalasini tanlash (rubrikalar)					

Biror oydinlashtirish kerakmi?

<u>Item #</u>		0	1	2	3	4
<u>56</u>	Alohida baholash maqsadida muvofiq bandlar yoki vazifalarni tanlash					
<u>57</u>	Reyting shkalalarini (rubrikalar) ishlatishni boshqalarga muvofiq ravishda o'rgatish					
<u>58</u>	Tilni baholash uchun boshqalarni yaxshi sifatli bandlar (savollar) yoki vazifalar yozishga o'rgatish					

<u>59</u>	Tilni baholash uchun yaxshi sifatli savollar yoki vasifalar yozish					
<u>60</u>	Testlarni tajribali tuzilmalarga (masalan: the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Language [ACTFL])					
<u>61</u>	O'tish yoki o'tolmaslik, yoki balni kamaytirishni aniqlash					
<u>62</u>	Baholashda yonbosishni aniqlash					
<u>63</u>	Baholash vazifasi uchun ball berish kalitlari va reyting shkalalarini (rubrikalar) loyhalashtirish					
<u>64</u>	Tilni qaysi aspektlarini baholash to'g'risida xulosa qilish					
<u>65</u>	Amalga oshirishdan oldin baholashlarni boshqarish/qo'llab ko'rish					
<u>66</u>	Ishlatishga tayyor muvofiq baholashlarni tanlash					

Biror ovdinlashirish kerakmi?

3-Qism Biyografik Ma'lumot

JINSI: _____

YOSHI: _____

UNIVERSITETDAGI INGILIZ TILI O'QITUVCHISI SIFATIDA TAJRIBA

YILLARI: _____

BU TAHLIL UCHUN BIROR MASLAHATLAR BORMI?

Bu tahlilga oz xissangizni qo'shganligingiz uchun katta rahmat!

Часть 1

На сколько, по вашему мнению, учителя должны обладать знаниями о каждом аспекте оценивании знаний о языке, приведенном ниже? Пожалуйста, ответьте в соответствии со следующей шкалой:

0 = учитель может ничего не знать о данном аспекте

1 = учителю можно знать немного о данном аспекте

2 = учитель может иметь поверхностные знания о данном аспекте

3 = учитель должен обладать хорошими знаниями о данном аспекте

4 = учитель должен обладать глубокими знаниями о данном аспекте

пункт		0	1	2	3	4
1	как использовать оценивание для информирования о целях обучения или изучения					
2	как использовать оценивание для определения прогресса в изучении языка					
3	как использовать оценивание для определения достижений в изучении языка					
4	как использовать оценивание для определения сильных и слабых сторон учащихся					

5	как использовать оценивание для мотивации учащихся					
6	как использовать само оценивание и оценивание в парах					
7	как правильно интерпретировать результаты оценивания					
8	как интерпретировать погрешность измерения					
9	как распознать что определенная оценка говорит о языковых способностях кого-либо					
<u>Нужны ли Вам какие-то разъяснения?</u>						
пункт		0	1	2	3	4
10	как определить совпадают ли стандарты оценивания с образовательной системой					
11	как определить соответствует ли содержание оценивания уровня владения языком культурным нормам					
12	как определить соответствуют ли результаты оценивания местному контексту					
13	как объяснять результаты и решения об оценивании учителям					
14	как объяснять результаты и решения об оценивании учащимся и их родителям					
15	как определить является ли оценивание уровня владения языком полезным для определенной цели					
16	как узнать что оценивание используется неправильно					
17	как подготовить учащихся к оцениванию уровня владения языком					
<u>Нужны ли Вам какие-то разъяснения?</u>						

пункт		0	1	2	3	4
18	как найти информацию, которая помогла бы в интерпретации результатов оценивания					
19	как давать полезные советы основываясь на результатах оценивания					
20	как оценивание может повлиять на структуру языкового курса или учебный план					
21	как оценивание может повлиять на учебные материалы					
22	как оценивание может повлиять на учебный процесс на уроке					
23	как развиваются языковые навыки					
24	как изучаются иностранные языки					
25	как язык используется в обществе					
26	как общественные ценности могут повлиять на использование языкового оценивания					
<u>Нужны ли Вам какие-то разъяснения?</u>						
Пункт		0	1	2	3	4
27	как устанавливаются проходные баллы и оценивающий срез					
28	концепт надежности оценивания					
29	концепт обоснованности оценивания					
30	структура языка					
31	преимущества и недостатки стандартных тестов					
32	философия дизайна определенного оценивания уровня владения языком					

33	влияние, которое может оказать оценивание уровня владения языком на общество					
34	соответствующие правовые нормы для оценивания в данной окрестности					
35	традиции оценивания в местном контексте					
Нужны ли Вам какие-то разъяснения?						
		0	1	2	3	4
36	специальная терминология, относящаяся к оцениванию уровня владения языком					
37	различные системы владения языком, например «Общеввропейские компетенции владения иностранным языком: изучение, преподавание, оценка», «Американский совет по преподаванию иностранных языков» (CEFR, ACTFL)					
38	различные стадии владения языком					
39	различные виды целей для оценивания уровня владением языком					
40	различные формы альтернативного оценивания (например, портфолио)					
41	свои личные убеждения по отношению к оцениванию уровня владения языком					
42	как свои личные убеждения могут повлиять на его практику оценивания					
43	как свои личные убеждения могут быть в конфликте с интересами других групп вовлеченными в оценивание					
44	как свои личные знания по оцениванию уровня					

	владения языком могут быть больше углублены					
<u>Нужны ли Вам какие-то разъяснения?</u>						

Часть II

На сколько, по вашему мнению, люди в выбранной вами профессии должны обладать навыками о каждом аспекте оценивании знаний о языке, приведенном ниже? Пожалуйста, ответьте в соответствии со следующей шкалой:

0 = может ничего не знать о данном аспекте
 1 = может знать немного о данном аспекте
 2 = может иметь поверхностные знания о данном аспекте
 3 = должен обладать хорошими знаниями о данном аспекте
 4 = должен обладать глубокими знаниями о данном аспекте

<u>Пункт</u>		0	1	2	3	4
45	использование статистики для анализа уровня сложности каждого вопроса					
46	использование статистики для анализа общей оценки определенного оценивания					
47	использование статистики для анализа качества определенных вопросов/заданий					
48	использование других методов кроме статистических (опросники, интервью, языковой анализ) для сбора информации о качестве оценивания уровня владения языком					
49	использование рейтинговые шкалы для оценивания говорения или письма					

50	использование характеристик для написания вопросов и заданий					
51	оценивание вопросов с закрытым ответом (например, вопрос с множественным ответом)					
52	оценивание вопросов с открытым ответом (например, вопрос с односложным ответом)					
53	разработка оценивания основанного на портфолио					
54	разработка характеристик (общий план) для оценивания уровня владения языком					
55	подборка подходящих рейтинговых шкал (рубрик)					
<u>Нужны ли Вам какие-то разъяснения?</u>						
<u>Пункт</u>		0	1	2	3	4
56	Подборка подходящего задания для определенной цели оценивания					
57	Подготовка других людей по использованию рейтинговых шкал (рубрик)					
58	Подготовка других людей для написания вопросов или заданий хорошего качества для оценивания уровня владения языком					
59	Написание вопросов и заданий хорошего качества для оценивания уровня владения языком					
60	Выверка тестов с различными системами владения языком, например «Общеввропейские компетенции владения иностранным языком: изучение, преподавание, оценка», «Американский					

	совет по преподаванию иностранных языков» (CEFR, ACTFL)					
61	Определение проходных баллов и оценивающих срезов					
62	Определение необъективности оценивания					
63	Разработка ключей оценивания и рейтинговых шкал (рубрик) для оценивающих заданий					
64	Принятие решений о том, какие аспекты языка оценивать					
65	Пилотирование/апробация оценивающих заданий до их проведения					
66	Подборка подходящих готовых оценивающих документов					
<u>Нужны ли Вам какие-то разъяснения?</u>						

Часть III – Личная информация

ПОЛ: _____

ВОЗРАСТ: _____

ОПЫТ РАБОТЫ УЧИТЕЛЕМ АНГЛИЙСКОГО ЯЗЫКА В УНИВЕРСИТЕТЕ (СКОЛЬКО ЛЕТ): _____

ЕСТЬ ЛИ У ВАС КАКИЕ-ЛИБО ПРЕДЛОЖЕНИЯ ДЛЯ ДАННОГО ОПРОСНИКА?

Appendix E: Uzbek Pseudonyms

<u>Focus Group</u>	<u>Teacher Number</u>	<u>Pseudonym</u>
1	1	Aisara
1	2	Klara
1	3	Ajva
1	4	Albina
1	5	Kamara
1	6	Leila
1	7	Svetlana
1	8	Lena
1	9	Shaknoza
1	10	Ulugbek
2	1	Diora
2	2	Almina
2	3	Ashura
2	4	Diliya
2	5	Gabriella
2	6	Abdulaziz
2	7	Umida
2	8	Guldasta
2	9	Madina
2	10	Maimouna
2	11	Nilufar
3	1	Bibidana
3	2	Darisa
3	3	Durdona
3	4	Madison
3	5	Farhod
3	6	Ona
3	7	Olmar
3	8	Nodira
3	9	Nargiza
3	10	Aziza
3	11	Kamila
4	1	Zumara
4	2	Samia
4	3	Mohira
4	4	Yulduz
4	5	Tahmina
4	6	Feruza
4	7	Zamifra
4	8	Zulnara
4	9	Farrukh
4	10	Shaholo

4	11	Sakina
5	1	Eldman
5	2	Nozliya
5	3	Aida
5	4	Chinara
5	5	Dana
5	6	Alsu
5	7	Mukaddas
5	8	Parizoda
5	9	Nozima
5	10	Yayra

Appendix F: Focus Group Protocol

Guiding Questions

Thank you everyone for coming to this focus group. I really appreciate it and look forward to getting to know each of you more. Today, I would like to know more about your teaching situations (contexts), your students, the type of decisions you should make in the classroom, and how you assess/test students. Let's first go around and give a brief self-introduction in English.

[Teacher's self-introduction – name, region in Uzbekistan, and the names of the classes he or she teaches]

Thank you everyone for your self-introductions. You all come from such diverse areas throughout the country of Uzbekistan.

- (1) Do they have any kind of set curriculum or can you decide for yourself what you will teach and assess?
- (2) What are your learning goals for their students, and what are your students' goals? Do these goals ever conflict?

- (3) In your classes, how are the grades determined for each student?
- (4) Have you had any training in testing/assessment?
- (5) At each of your universities/colleges, how do you track student's (spoken, written, listening, reading, grammar, etc.) language learning?
- (6) How do you decide what to test?
- (7) How do you make tests?
- (8) What do you do with the results of the tests?
- (9) If a student does not pass a test or class, what are some consequences?
- (10) Describe a challenge you have had in assessing your students?

Appendix G: Semi-Structured Interview Protocol

Individual Interview

Interviewer self-introduction: Good morning/ afternoon/ evening Mrs./Mr. XXX, my name is David Chiesa and I am the EL Specialist and also a student at Georgia State University in the United States. Thank you very much for coming today to talk to me more about your specific experiences with language assessment.

The study I am conducting is looking at how social, cultural, economic, and political factors influence what and how teachers think about and do with assessment. I would like to talk to you to understand more about your life experiences.

Before we begin, can you tell me a little bit about yourself? Where do you currently teach? (Please point it out on a map to me.) How long have you been teaching English? What classes are you currently teaching?

Thank you very much. Let's begin.

Possible Guiding Questions:

- A. Can you please tell me about a time in your life / an experience when you took a high stakes language test? How did you feel?"
- B. Can you tell me about an interesting experience when you were trying to assess a student's language abilities?

- C. Can you remember how you felt when YOU were being assessed/tested?
- D. Are you respected more by your peers if your students do well on a test?" How is language testing looked at by your administration, at your university?
- E. Do students get/receive more money or prestige if they do well on language tests?" What happens if they do not do well?
- F. What are the economic benefits to students of doing well in school?
- G. How large a role does the department head, dean, or higher-up administrative person have in the creation, implementation, and scoring of your language tests?
- H. Can you talk about/discuss any Uzbekistani cultural rules (spoken and not spoken) for assessing students?

Appendix H: Descriptive Statistics for Each Item

<u>Item</u>	<u>Item Description</u>	<u>Mean</u>	<u>Median</u>	<u>Mode</u>	<u>SD</u>	<u>Min</u>	<u>Max</u>	<u>Range</u>
23	<i>how language skills develop</i>	3.23	3.00	3.00	0.76	1.00	4.00	3.00
22	<i>how assessments can influence teaching and learning in the classroom</i>	3.09	3.00	3.00	0.81	1.00	4.00	3.00
21	<i>how assessments can influence teaching and learning materials</i>	3.06	3.00	3.00	0.78	1.00	4.00	3.00
19	<i>how to give useful feedback on the basis of assessment</i>	3.05	3.00	3.00	0.80	1.00	4.00	3.00
3	<i>how to use assessments to evaluate achievement in language learning</i>	3.04	3.00	3.00	0.78	0.00	4.00	4.00
30	<i>the structure of language</i>	3.03	3.00	3.00	0.84	0.00	4.00	4.00
24	<i>how foreign/second languages are learned</i>	2.98	3.00	3.00	0.88	1.00	4.00	3.00
58	<i>training others to write good quality items (questions) or tasks for language assessment</i>	2.97	3.00	3.00	0.86	0.00	4.00	4.00
37	<i>different language proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council</i>	2.95	3.00	3.00	0.87	0.00	4.00	4.00

	<i>on the Teaching of Foreign Languages [ACTFL])</i>							
5	<i>how to use assessments to motivate student learning</i>	2.93	3.00	3.00	0.94	0.00	4.00	4.00
39	<i>different types and purposes for language assessment (e.g., proficiency, achievement, diagnostic)</i>	2.91	3.00	3.00	0.83	1.00	4.00	3.00
17	<i>how to prepare learners to take language assessments</i>	2.90	3.00	3.00	0.94	0.00	4.00	4.00
18	<i>how to find information to help in interpreting results</i>	2.89	3.00	3.00	0.88	1.00	4.00	3.00
1	<i>how to use assessments to inform learning or teaching goals</i>	2.87	3.00	2.00	0.89	1.00	4.00	3.00
38	<i>different stages of language proficiency</i>	2.86	3.00	3.00	0.87	0.00	4.00	4.00
60	<i>aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference)</i>	2.86	3.00	3.00	0.90	0.00	4.00	4.00
6	<i>how to use self- and peer-assessment</i>	2.85	3.00	3.00	0.88	0.00	4.00	4.00
20	<i>how assessments can influence the design of a language course or curriculum</i>	2.85	3.00	3.00	0.78	1.00	4.00	3.00
9	<i>how to interpret what a particular score says about an individual's language ability</i>	2.85	3.00	3.00	0.81	1.00	4.00	3.00
15	<i>how to determine whether a language assessment is useful for a particular purpose.</i>	2.85	3.00	3.00	0.91	0.00	4.00	4.00
50	<i>using specifications to develop items (questions) and tasks</i>	2.85	3.00	3.00	0.84	1.00	4.00	3.00
64	<i>making decisions about what aspects of language to assess</i>	2.85	3.00	3.00	0.95	0.00	4.00	4.00
26	<i>how social values can influence language assessment design and use</i>	2.84	3.00	3.00	0.86	0.00	4.00	4.00
54	<i>developing specifications (overall plans) for language assessment</i>	2.84	3.00	3.00	0.82	0.00	4.00	4.00
4	<i>how to use assessments to diagnose learners' strengths and weaknesses</i>	2.85	3.00	3.00	0.95	0.00	4.00	4.00

48	<i>using techniques other than statistics (e.g., questionnaires, interviews, analysis of language) to get information about the quality of language assessment</i>	2.82	3.00	3.00	0.87	1.00	4.00	3.00
51	<i>scoring closed-response questions (e.g., Multiple Choice Questions)</i>	2.82	3.00	3.00	0.85	1.00	4.00	3.00
57	<i>training others to use rating scales (rubrics) appropriately</i>	2.82	3.00	3.00	0.87	0.00	4.00	4.00
59	<i>writing good quality items (questions) or tasks for language assessments</i>	2.82	3.00	3.00	1.00	0.00	4.00	4.00
49	<i>using rating scales to score speaking or writing performances</i>	2.79	3.00	3.00	0.85	0.00	4.00	4.00
65	<i>piloting/trying-out assessments before their administration</i>	2.79	3.00	3.00	0.95	0.00	4.00	4.00
13	<i>how to communicate assessment results and decisions to teachers</i>	2.78	3.00	3.00	0.94	0.00	4.00	4.00
52	<i>scoring open-ended questions (e.g., short answer questions)</i>	2.78	3.00	3.00	0.87	1.00	4.00	3.00
63	<i>designing scoring keys and rating scales (rubrics) for assessment tasks</i>	2.78	3.00	3.00	0.91	0.00	4.00	4.00
33	<i>the impact language assessments can have on society</i>	2.77	3.00	3.00	0.97	0.00	4.00	4.00
29	<i>the concept of validity</i>	2.77	3.00	3.00	1.01	0.00	4.00	4.00
2	<i>how to use assessments to evaluate progress in language learning</i>	2.76	3.00	3.00	0.81	1.00	4.00	3.00
36	<i>the specialist terminology related to language assessment.</i>	2.76	3.00	3.00	0.94	0.00	4.00	4.00
66	<i>selecting appropriate ready-made assessments.</i>	2.76	3.00	3.00	0.96	1.00	4.00	3.00
25	<i>how language is used in society</i>	2.75	3.00	3.00	0.94	0.00	4.00	4.00
53	<i>developing portfolio-based assessments</i>	2.75	3.00	3.00	0.86	0.00	4.00	4.00
7	<i>how to interpret assessment results appropriately</i>	2.74	3.00	3.00	0.84	1.00	4.00	3.00
28	<i>the concept of reliability</i>	2.74	3.00	3.00	0.91	0.00	4.00	4.00
62	<i>identifying assessment bias</i>	2.73	3.00	3.00	1.00	0.00	4.00	4.00

61	<i>determining pass-fail marks or cut-scores</i>	2.72	3.00	3.00	0.82	1.00	4.00	3.00
8	<i>how to interpret measurement error</i>	2.71	3.00	2.00	0.89	0.00	4.00	4.00
40	<i>different forms of alternative assessments (e.g., portfolio assessment)</i>	2.70	3.00	3.00	0.93	0.00	4.00	4.00
12	<i>how to determine if the results of the assessment are relevant to the local context</i>	2.67	3.00	3.00	0.83	1.00	4.00	3.00
44	<i>how one's own knowledge of language assessment might be further developed</i>	2.66	3.00	3.00	1.08	0.00	4.00	4.00
55	<i>selecting appropriate rating scales (rubrics)</i>	2.65	3.00	3.00	0.92	0.00	4.00	4.00
11	<i>how to determine if the content of a language assessment is culturally appropriate</i>	2.65	3.00	3.00	0.82	1.00	4.00	3.00
42	<i>how one's own beliefs/attitudes might influence one's assessment practices</i>	2.64	3.00	3.00	1.04	0.00	4.00	4.00
56	<i>selecting appropriate items or tasks for a particular assessment purpose</i>	2.64	3.00	3.00	0.98	0.00	4.00	4.00
16	<i>how to recognize when an assessment is being used inappropriately</i>	2.64	3.00	3.00	1.00	0.00	4.00	4.00
32	<i>the philosophy behind the design of a relevant language assessment</i>	2.62	3.00	3.00	0.93	0.00	4.00	4.00
31	<i>the advantages and disadvantages of standardized testing</i>	2.62	3.00	3.00	0.97	0.00	4.00	4.00
47	<i>using statistics to analyze the quality of individual items (questions)/tasks</i>	2.61	3.00	3.00	0.88	0.00	4.00	4.00
35	<i>the assessment traditions in a local context</i>	2.58	2.00	2.00	1.04	0.00	4.00	4.00
34	<i>the relevant legal regulations for assessment in the local area</i>	2.58	3.00	3.00	1.00	0.00	4.00	4.00
10	<i>how to determine if a local assessment aligns with a local education system</i>	2.57	3.00	3.00	0.87	0.00	4.00	4.00
27	<i>how pass-fail marks or cut-off scores are set</i>	2.56	3.00	0.00	0.91	0.00	4.00	4.00
41	<i>one's own beliefs/attitudes towards language assessment</i>	2.55	3.00	3.00	0.91	0.00	4.00	4.00

46	<i>using statistics to analyze overall scores on a particular assessment</i>	2.53	2.00	2.00	0.95	0.00	4.00	4.00
45	<i>using statistics to analyze the difficulty of individual items (questions)</i>	2.52	2.00	2.00	1.00	0.00	4.00	4.00
14	<i>how to communicate assessment results and decisions to students or parents</i>	2.52	2.00	2.00	0.93	0.00	4.00	4.00
43	<i>how one's own beliefs/attitudes may conflict with those of other groups involved in assessment</i>	2.35	2.00	3.00	1.07	0.00	4.00	4.00

Appendix I: Operationalization of LAL Constructs

Taylor (2013) Framework	Language Testing Journal (2013)	Harding and Kremmel Poster (2016)
Knowledge of Theory	<p>“...learning theories and practices and evolving theories of language and culture” (Scarino, p. 313).</p> <p>“...It is important for teachers to have a sense of expected norms as levels of achievement at different phases along the continuum of learning” (Scarino, p. 314).</p>	Theoretical Knowledge About Language and Language Learning
Technical Skills	<p>“... constructing and evaluating language tests” (Scarino, p. 314).</p> <p>“... Davies (2008) ... skills (the how-to- or basic testing expertise” (Malone, p. 330).</p>	Language Assessment Construction; Language Assessment Administration/ Scoring
Principles and Concepts	<p>Contrasting paradigms – traditional and alternative assessment – “Teachers of languages need to understand the assumptions of both paradigms and move between them” (Scarino, p. 312).</p> <p>“A general understanding of assessments as compared to tests and formative versus summative tests; knowledge of different types of language assessments and what information each type provides...”</p>	Principle and Concepts

“...Davies (2008) Principles (concepts underlying testing such as validity, reliability, and ethics)” (Malone, p. 331).

“... Having the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results” (Jeong, p. 346).

Language
Pedagogy

“... Furthermore, in school language education, assessment cannot be separated from its relationship with the curriculum, and processes of teaching and learning” (Scarino, p. 314).

Language Pedagogy

“It [assessment] can and should integrate with teaching, forming a relationship in which assessment informs and improves teaching and vice versa” (Malone, p. 330).

Sociocultural
Values

The social turn in language assessment... “...One module needs to provide background to the social, educational and political aspects of assessment... as well as the critical views on the role of language tests in society” (Scarino, 314).

Sociocultural
Values

“The concept of language assessment literacy is underpinned by a view of language assessment as a social practice rather than simply as a technical activity... a critical understanding about the roles and functions of assessment within education and society” (O’Loughlin, p. 363).

Local Practices

Contextual Consideration: “Assessment is always situated in distinctive institutional and policy contexts that confer on the assessment process particular characteristics and requirements” (Scarino, p. 311) – these institutional and policy requirements create a culture of certainty and compliance that is not easily challenged by teachers.

Local Practices

Personal
beliefs/attitudes

“A good understanding of the contrasting paradigms provides teachers with a basis for understanding the tensions they often experience in practice” (Scarino, p. 313).

Beliefs and
Attitudes

“More than just having a knowledge base, they need to understand and question the assumptions made and the possibilities and limitations of the different paradigms” (Scarino, p. 313)

“... requires a consideration of the processes for its development that invite teachers to examine, in a critical way, their own conceptions of the assessment process itself and the conceptions of others” (Scarino, p.314)

Scores and
Decision
Making

“...interpret and analyze assessment results, respond appropriately to the results and their meanings, and use the results in their teaching” (p. 331).

Scores and Decision
Making

“AL also incorporates teachers’ ability to communicate assessment results effectively to students, parents, and other educational professionals (Stiggins, 1999); (Jeong, 2013) The interpretation and use of proficiency test scores for university administration. (test-users)

“... understand the risk of making decisions about the fate of human beings using fallible language tests” (O’Loughlin, p. 365).

Appendix J: Kremmel and Harding, LAL Operationalization of Constructs



Appendix K: Researcher's Finalized Coding Scheme

Item	Item Description	Code
1	<i>how to use assessments to inform learning or teaching goals</i>	ScDeMa
2	<i>how to use assessments to evaluate progress in language learning</i>	ScDeMa
3	<i>how to use assessments to evaluate achievement in language learning</i>	ScDeMa
4	<i>how to use assessments to diagnose learners' strengths and weaknesses</i>	ScDeMa
5	<i>how to use assessments to motivate student learning</i>	ScDeMa
6	<i>how to use self- and peer-assessment</i>	ScDeMa
7	<i>how to interpret assessment results appropriately</i>	ScDeMa
8	<i>how to interpret measurement error</i>	ScDeMa
9	<i>how to interpret what a particular score says about an individual's language ability</i>	ScDeMa
10	<i>how to determine if a local assessment aligns with a local education system</i>	LocPrac
11	<i>how to determine if the content of a language assessment is culturally appropriate</i>	SocVal
12	<i>how to determine if the results of the assessment are relevant to the local context</i>	LocPrac
13	<i>how to communicate assessment results and decisions to teachers</i>	ScDeMa
14	<i>how to communicate assessment results and decisions to students or parents</i>	ScDeMa
15	<i>how to determine whether a language assessment is useful for a particular purpose</i>	PrCon
16	<i>how to recognize when an assessment is being used inappropriately</i>	ScDeMa
17	<i>how to prepare learners to take language assessments</i>	LanPed
18	<i>how to find information to help in interpreting results</i>	ScDeMa
19	<i>how to give useful feedback on the basis of assessment</i>	ScDeMa
20	<i>how assessments can influence the design of a language course or curriculum</i>	LanPed
21	<i>how assessments can influence teaching and learning materials</i>	LanPed
22	<i>how assessments can influence teaching and learning in the classroom</i>	LanPed
23	<i>how language skills develop</i>	KnOfTh
24	<i>how foreign/second languages are learned</i>	KnOfTh
25	<i>how language is used in society</i>	SocVal
26	<i>how social values can influence language assessment design and use</i>	SocVal
27	<i>how pass-fail marks or cut-scores are set</i>	PrCon
28	<i>the concept of reliability</i>	PrCon
29	<i>the concept of validity</i>	PrCon
30	<i>the structure of language</i>	KnOfTh
31	<i>the advantages and disadvantages of standardized testing</i>	PerBelAtt
32	<i>the philosophy behind the design of a relevant language assessment</i>	PrCon
33	<i>the impact language assessments can have on society</i>	SocVal
34	<i>the relevant legal regulations for assessment in the local area</i>	LocPrac
35	<i>the assessment traditions in a local context</i>	LocPrac
36	<i>the specialist terminology related to language assessment</i>	PrCon

37	<i>different language proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])</i>	KnOfTh
38	<i>different stages of language proficiency</i>	KnOfTh
39	<i>different types of purposes for language assessment purposes (e.g., proficiency, achievement, diagnostic)</i>	PrCon
40	<i>different forms of alternative assessments (e.g., portfolio assessment)</i>	PrCon
41	<i>one's own beliefs/attitudes towards language assessment.</i>	PerBelAtt
42	<i>how one's own beliefs/attitudes might influence one's assessment practices</i>	PerBelAtt
43	<i>how one's own beliefs/attitudes may conflict with those of other groups involved in assessment</i>	PerBelAtt
44	<i>how one's own knowledge of language assessment might be further developed</i>	PerBelAtt
45	<i>using statistics to analyze the difficulty of individual items (questions)</i>	TeSk
46	<i>using statistics to analyze overall scores on a particular assessment</i>	TeSk
47	<i>using statistics to analyze the quality of individual items (questions)/tasks</i>	TeSk
48	<i>using techniques other than statistics (e.g., questionnaires, interviews, analysis of language) to get information about the quality of language assessment</i>	TeSk
49	<i>using rating scales to score speaking or writing performances</i>	TeSk
50	<i>using specifications to develop items (questions) and tasks</i>	TeSk
51	<i>scoring closed-response questions (e.g., Multiple Choice Questions)</i>	TeSk
52	<i>scoring open-ended questions (e.g., short answer questions)</i>	TeSk
53	<i>developing portfolio-based assessments</i>	TeSk
54	<i>developing specifications (overall plans) for language assessments</i>	TeSk
55	<i>selecting appropriate rating scales (rubrics)</i>	TeSk
56	<i>selecting appropriate items or tasks for a particular assessment purpose.</i>	PrCon
57	<i>training others to use rating scales (rubrics) appropriately</i>	TeSk
58	<i>training others to write good quality items (questions) or tasks for language assessment.</i>	TeSk
59	<i>writing good quality items (questions) or tasks for language assessments.</i>	TeSk
60	<i>aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])</i>	PrCon
61	<i>determining pass-fail marks or cut-scores</i>	TeSk
62	<i>identifying assessment bias.</i>	TeSk
63	<i>designing scoring keys and rating scales (rubrics) for assessment tasks.</i>	TeSk
64	<i>making decisions about what aspects of language to assess.</i>	PrCon
65	<i>piloting/trying-out assessments before their administration</i>	PrCon
66	<i>selecting appropriate ready-made assessments.</i>	PrCon

Appendix L: Summarizing Problematic Item Concerns

Item	Item Description	Dis	Am	Tran	K&H
1	<i>how to use assessments to inform learning or teaching goals</i>				X
2	<i>how to use assessments to evaluate progress in language learning</i>				X
3	<i>how to use assessments to evaluate achievement in language learning</i>				X
4	<i>how to use assessments to diagnose learners' strengths and weaknesses</i>				X
5	<i>how to use assessments to motivate student learning</i>				X
6	<i>how to use self- and peer-assessment</i>				X
8	<i>how to interpret measurement error</i>			X	
10	<i>how to determine if a local assessment aligns with a local education system</i>			X	
11	<i>how to determine if the content of a language assessment is culturally appropriate</i>	X			
12	<i>how to determine if the results of the assessment are relevant to the local context</i>			X	
13	<i>how to communicate assessment results and decisions to teachers</i>	X			
14	<i>how to communicate assessment results and decisions to students or parents</i>	X			
15	<i>how to determine whether a language assessment is useful for a particular purpose</i>		X		
16	<i>how to recognize when an assessment is being used inappropriately</i>				X
17	<i>how to prepare learners to take language assessments</i>			X	
18	<i>how to find information to help in interpreting results</i>	X			
19	<i>how to give useful feedback on the basis of assessment</i>	X			X
20	<i>how assessments can influence the design of a language course or curriculum</i>		X		
21	<i>how assessments can influence teaching and learning materials</i>		X		
22	<i>how assessments can influence teaching and learning in the classroom</i>		X		
27	<i>how pass-fail marks or cut-off scores are set</i>				X
31	<i>the advantages and disadvantages of standardized testing</i>		X		
32	<i>the philosophy behind the design of a relevant language assessment</i>		X		X

34	<i>the assessment traditions in a local context</i>		X	
35	<i>the relevant legal regulations for assessment in the local area</i>		X	
39	<i>different types of purposes for language assessment purposes (e.g., proficiency, achievement, diagnostic)</i>		X	
52	<i>how one's own knowledge of language assessment might be further developed</i>		X	
49	<i>using rating scales to score speaking or writing performances</i>		X	
51	<i>scoring closed-response questions (e.g., Multiple choice questions)</i>		X	
52	<i>scoring open-ended questions (e.g., short answer questions)</i>		X	
54	<i>developing specifications (overall plans for language assessments)</i>		X	
56	<i>selecting appropriate items or tasks for a particular assessment purpose.</i>	X		X
60	<i>aligning tests to proficiency frameworks (e.g., the Common European Framework of Reference [CEFR], American Council on the Teaching of Foreign Languages [ACTFL])</i>	X		X
61	<i>determining pass-fail marks or cut-scores</i>		X	
62	<i>identifying assessment bias</i>		X	X
66	<i>selecting appropriate ready-made assessments.</i>		X	X

Appendix M: Final Examination Example Sample, PRESETT

ENGLISH PHONETICS AND PHONOLOGY DEPARTMENT

FINAL TEST FOR THE 4th YEAR STUDENTS

Name _____

Teacher _____

Group _____

Sign _____

Date _____

Total point _____ Variant 1

Listening	Reading	Language structure	Writing
8	8	7	7

I LISTENING

1 Listen to five people talking about the disadvantages of various jobs. Match the speakers (1–5) to the disadvantages of the job that they mention (A–H).

- Speaker 1
 Speaker 2
 Speaker 3
 Speaker 4
 Speaker 5

- A lack of job security
 B the possibility of making a serious mistake
 C having to work at inconvenient times
 D bad treatment from employers
 E having to work in unpleasant physical conditions
 F other people's low opinion of the job
 G having to be dishonest
 H the need to cooperate with colleagues

	2
--	---

2 Listen to five people talking about various historical films. Match the speakers (1–5) to what they say about the films (A–H).

- Speaker 1
 Speaker 2
 Speaker 3
 Speaker 4
 Speaker 5

- A It was different from what I had expected before I saw it.
 B It has a personal connection for me.
 C Most people remember one particular scene from it.
 D I liked it so much that I saw it over and over again.
 E I remained affected by it for some time after I'd seen it.
 F One particular scene affected me emotionally a great deal.
 G The acting is the most impressive aspect of it.
 H I found elements of it unrealistic.

	2
--	---

3 Listen to two friends discussing an article about personality types. Tick (✓) A, B, or C.

- 1 The two speakers agree that _____.
 - A the man does not react well to pressure
 - B other people like the man's attitude and behaviour
 - C the man should not consider himself a 'go-getter'

- 2 The man doesn't agree that he _____.
 - A is good at organizing other people
 - B makes too little effort
 - C annoys other people

- 3 What do the two speakers agree on the subject of discussions?
 - A The man often changes his view during them.
 - B The man always wants to make other people agree with him.
 - C The man likes it when people disagree with him.

- 4 The woman agrees that she _____.
 - A is regarded as unreliable by many people
 - B has a relaxed attitude to life
 - C frequently changes her plans

- 5 The man says that one characteristic of 'performers' is that they _____.
 - A think too much about criticism
 - B expect too much of other people
 - C are too loyal to other people

	2
--	---

Listen to an interview about computer addiction. Tick (✓) A, B, or C.

- 1 What does the interviewer say about computer addiction in his introduction?
 - A It causes disagreement among experts.
 - B It affects a great many people.
 - C It is unlikely that it really exists.

- 2 Colin says that one reason why computer addiction is a difficult subject is that _____.
 - A it is easy for people to be addicted to computers without realizing it
 - B people don't want to think that it is similar to common addictions
 - C reasons why people use computers for a long time vary considerably

- 3 What is Colin's point about hobbies?
 - A Many people spend longer doing hobbies than using computers.
 - B People don't usually think that a hobby can be an addiction.
 - C They can be just as addictive as computer use.

- 4 What does Colin say about computer-game addiction?
- A Official bodies are likely to accept that it exists in the future.
- B It is something that many people might notice.
- C Experts cannot agree on a clear definition of it.
- 5 What is Colin's personal opinion on computer addiction?
- A Some people think they have an addiction but they don't.
- B More people will believe that computer use can become an addiction.
- C Certain behaviour probably indicates the existence of an addiction.

	2
--	---

Listening total	8
-----------------	---

II READING COMPREHENSION

Read the article and tick (✓) A, B, or C.

CREATING YOUR OWN JOB SATISFACTION

A lot of people think that job satisfaction is only for other people. They look with envy on people who love their jobs and who don't think of them as work. The idea of earning a living from something you really like doing only applies to a few very lucky people, they think. But this isn't so. It's not only people in the so-called 'glamour professions', for example, who can get genuine job satisfaction. You don't have to be in the arts or a sports person to get enjoyment from what you do for a living. Even if you're in a boring job, it's quite possible to get some satisfaction from it.

The key to this is your attitude. You may think it's unlikely that you can derive much satisfaction from a dull job that doesn't require much thought and that involves a lot of routine procedures. But if you approach it with the right attitude, and put some effort in, you may be surprised at how enjoyable you can make it. Of course, if you just sit there telling yourself how boring your job is, you'll never get anything out of it. But if you set out to find ways of making it enjoyable, there's a good chance you'll manage to.

One thing you can do is to set yourself challenges. Think about what you can do for yourself to make your work a little bit more interesting. If you've got a repetitive job, set yourself some targets to meet and try to beat your previous records. Or use your initiative in other ways. Think about ways you could develop your career into more interesting areas – see how you could improve your skills by doing a course, for example, or look into new skills you could get that would stand you in good stead for the future.

For some people, it's not boredom that's the problem, it's the fact that their jobs involve a significant amount of unpleasantness. But if you keep telling yourself your job is horrible and there's nothing you can do about it, you'll get stuck in a rut and you'll never get out of it. Focus on developing a positive attitude and try to keep any negative thoughts about your job out of your mind. Keep a sense of perspective – if something's gone wrong on a particular day, decide whether it really matters or not. If it isn't actually all

that important, don't dwell on it, let it go. Learn from it, and take an upbeat view of what's happened – you'll know how to avoid the problem in future, or what to do about it if it happens again.

Of course, you may reach the point where you feel totally trapped in a job. If this mindset starts to overwhelm you, check out the options you may have in the place where you work. Map out a plan for improving your situation there. Are there other roles in the organization you could apply for? Could you ask to be given different tasks? Could you get a different kind of assignment or go to another department where your skills are required? Finding out there are other options will give you a sense of control over your own working life. Even if you have no job satisfaction right now, you'll feel better if you know that there is a realistic prospect of work that gives you a degree of it.

Expectations are another key ingredient in job satisfaction. Take a long hard look at what you really are capable of. Sometimes it's important to accept that you couldn't really do the much more interesting or high-powered job you aspire to. Try to be aware of your own strengths and weaknesses. Focus on the things you really are good at, rather than on things that, if you're really honest with yourself, you couldn't actually do. Make the most of your situation and feel good about what you can do, rather than feeling bad about what you can't do.

Work is a very important part of most people's lives and it's important to get at least some satisfaction from it. If you really dislike your time at work, the rest of your life is affected too, and you can easily get a negative outlook on life in general. It's in your own hands to avoid this. Even if you can't get the job of your dreams, you can take steps to create your own job satisfaction.

- 1 The writer's aim in the first paragraph is to _____.
 A distinguish between different kinds of work B correct a false belief
 C define the term 'job satisfaction'
- 2 In the second paragraph, the writer _____.
 A warns readers against a certain attitude B advises readers not to have unrealistic ideas about work C suggests to readers that most jobs are dull
- 3 In the third paragraph, the writer emphasizes _____.
 A how easy it can be to progress in a career
 B a particular route to job satisfaction
 C the need for people to motivate themselves
- 4 The writer uses the phrase 'stuck in a rut' to refer to _____.
 A a situation that won't improve B the bad behaviour of others
 C the danger of being too sensitive
- 5 What is the writer's advice if something goes wrong?
 A Pretend that it didn't happen. B Use the experience to your advantage.
 C Don't think about it until later.
- 6 The writer says that people who feel 'trapped' in a job should consider _____.
 A discussing their unhappiness with managers
 B moving to a different company or organization
 C changing the kind of work they do
- 7 The writer says that you will feel better about your working life if you _____.
 A think that some job satisfaction is possible in the future
 B choose a particular career option for the future
 C stop aiming for job satisfaction for a while

- 8 What does the writer say about expectations of work?
 A Too many people have unrealistic ones.
 B They should be based on real ability. C They may change as time passes.
- 9 In the final paragraph, the writer says that _____.
 A lack of job satisfaction has serious consequences for people
 B a negative attitude to life leads to a negative attitude to work
 C there is more to life than job satisfaction
- 10 Which of the following best sums up the writer's view in the text as a whole?
 A Some people find it easier to get job satisfaction than others.
 B Everyone can get a certain amount of job satisfaction.
 C Job satisfaction is the most important issue in the workplace today.

Reading

8

III Language structure

A. Grammar

1 Not only late, you're also not dressed properly.

- a) you are b) are you c) you're d) you aren't

2 If the bodyof the balanced nutrition it requires, dieting can be harmful.

- a) is deprived b) deprives c) deprived d) which deprives

3 It was wrong of you to interfere in her private affairs. Youhave imposed your wishes on her.

- a) mustn't b) might not c) shouldn't d) would not

4she could not read or write , she could retain a long list of names and addresses in her mind.

- a) Since b) Although c) In spite of d) Even

5 The classes must become smaller if English effectively.

- a) is to teach b) teaches c) has taught d) is to be taught

6 In many novels of the nineteenth century, the poor are depicted as being evil, dirty, and criminal., the rich are depicted as being kind, generous and virtuous.

- a) Conversely b) On the contrary c) Moreover d) Likewise

7 Thanks for your help.you want more information, call again tomorrow.

- a) Did b) would c) might d) should

B. Vocabulary competence

Complete the words in the sentences.

- 1 My mother was **b**_____ up in a small village in the mountains.
- 2 We were **e**_____ in conversation, we didn't notice that the restaurant had closed.
- 3 The person who earns most of the money to pay a family's expenses is the **b**_____.
- 4 I have a positive **o**_____ on life and I don't worry too much.
- 5 Something that is a bit different or unexpected can be described as **q**_____.
- 6 A journey to and from work is called **a**_____ .
- 7 A person who has been injured in a war is called a **c**_____ .

Language Structure		7
--------------------	--	---

IV WRITING

Write an article of approximately 250 words for an English-language magazine about how the area given below has changed in the last 20 years in your country and say whether you think the changes are positive or negative.

People's work-life balance

Writing		7
---------	--	---

Appendix N: National Standards – Higher Educational Levels

Appendix 1 to the Decree #____ of the
Cabinet of Ministries of the RUz
dated _____

UZBEKISTAN STATE STANDARD

State Educational Standards of Continuous Education of Uzbekistan

REQUIREMENTS

necessary for content and level of learners on foreign languages

TASHKENT – 2013

EXPLANATORY NOTE

State Education standards define the proficiency level of learners in foreign languages and serve as a basis for developing course syllabi, coursebooks, handbooks, regulations and other documents.

The core legislative acts for developing state educational standards on foreign languages are: the Constitution of the Republic of Uzbekistan, UNO Convention on children's rights, laws of the Republic of Uzbekistan "On Education", "On national programme for personnel training"; The Decree of the President of the Republic of Uzbekistan dated 10th December 2012 #PP-1875 "On measures to further improve foreign language learning system", decrees of the Cabinet of Ministries of the Republic of Uzbekistan "On approving State educational standards for public secondary education" № 390 dated 16 August 1999, «On approving State educational standards for secondary special vocational education» № 400 dated 16 October 2000; as well as

- O`z DSt 1.0:98. The System for state standardization of the Republic of Uzbekistan;
- O`z DSt 1.1-92. The System for state standardization of the Republic of Uzbekistan. The procedure for the development, coordination, approval and registration of the standards of the Republic of Uzbekistan;

- O`z DSt 1.9:1995 System of state standardization of the Republic of Uzbekistan; Procedure for developing, agreement, confirmation and registration of standards of the Republic of Uzbekistan;

- O`z DSt 6.38-90. The unified system of documents. The system of organisational-directive documents. Requirements to the documentation.

Foreign language learning in the Uzbekistan system of continuous education based on State educational standards is organized in the order given below:

STAGE OF EDUCATION	REQUIREMENTS TO GRADUATES	CEFR LEVELS	NAME OF THE LEVEL
Public secondary education	Primary school graduates	A 1	Basic user initial level
	Graduates of 9 th grade of public schools	A 2	Basic user level
	Graduates of 9 th grades in public specialized schools majoring in learning foreign languages	A 2+	Basic user enhanced level
Secondary special vocational education	Graduates of academic lyceums with non-language profile	B 1	Independent user initial level
	Graduates of vocational colleges		
	Graduates of academic lyceums with language profile – second foreign language	B 1+	Independent user high initial level
Higher Education	Graduates of bachelor’s degree courses in non-language departments of HEIs.	B 2	Independent user level
	Graduates of master’s degree courses in non-language departments of HEIs		
	Graduates of bachelor’s degree courses in language departments of HEIs - second foreign language		
	Graduates of bachelor’s degree courses in language departments of HEIs	C1	Proficient user level
	Graduates of master’s degree courses in language departments of HEIs		
Postgraduate education	Institute of senior researchers, non-language profile	B 2	Independent user level
	Institute of senior researchers, language profile	C1	Proficient user level

Based on the present standard benchmarks for state attestation, curricula and syllabi on foreign languages (English, French, German and others) are to be developed taking into account features of an educational establishment and approved by relevant resolutions of the Ministries.

THE STATE STANDARD STRUCTURE

The state standards on foreign languages for all levels of education include:

- ✓ aims and objectives of the academic course;
- ✓ content of teaching and learning;
- ✓ requirements for the exit level of graduates of educational establishments at all stages of education.

AIMS AND OBJECTIVES OF THE COURSE

The aim of teaching foreign language at all levels of education of the Republic of Uzbekistan is to develop the FL communicative competence of learners to be able to function in the multicultural world in everyday, academic and professional spheres.

Communicative competence in foreign language is generally defined as ‘both the tacit knowledge of a language and the ability to use it’.

It is commonly accepted that various important sub-competences contribute to overall communicative competence. For the purposes of this document, these competences are grouped together as follows:

Linguistic competence, which refers to knowledge of language areas (phonetics, vocabulary, grammar) and language skills (listening, speaking, reading and writing).

Sociolinguistic competence enables FL learners to choose appropriate linguistic forms, ways of expression depending on the setting, communicative aim and intentions of the speaker. Sociolinguistic competence also embraces socio-cultural competence, which refers to the ability to identify and respond appropriately to situations in the target culture and with speakers of the target language.

Pragmatic Competence refers to the ability to respond spontaneously to situations in the target language as they develop, and includes the deployment of strategies to take part effectively in interactions – for example, interrupting, clarifying, compensating when communication breaks down, etc. Discourse competence is subsumed under pragmatic competence for the purposes of this document. It refers to the ability to chain together idea with appropriate language in speech or in writing. It also assumes the ability to understand linguistic signals in connected speech or writing, for example sequencing devices, language to make contrast, ways of beginning and ending, etc.

CONTENT OF TEACHING AND LEARNING

The content of teaching and learning is presented in the form of themes, which are subject to compulsory inclusion in major educational programmes of public secondary, secondary special, graduate and postgraduate education. Learning material across all levels of education ensures continuity, consistency and recurrence in learning and teaching.

The Content section of the standards at each level may be used as a minimum basis for syllabus design and for textbook development.

REQUIREMENTS FOR THE EXIT LEVEL OF GRADUATES IN EACH SECTOR OF EDUCATION

Requirements for the exit level in foreign languages are developed in accordance with the required minimum content, are consistent across levels of public secondary, secondary special, graduate and postgraduate education and are presented in the form of ‘can do’ descriptors in the language skills, with supplementary guidance on grammar, vocabulary, phonology and spelling where required. The language skill descriptors are linked to and derived from those offered in the Common European Framework of Reference to ensure compatibility with international standards. They are described in simple, comprehensible terms as they need to be understood by learners as well as teachers and other stakeholders and interest groups. They should be seen as a way of describing **exit levels**, which means:

- i. that syllabus designers and textbook writers need to take account of the level(s) below when developing programmes and materials aimed at reaching the exit level.
- ii. they are intended for the purpose of developing assessment tools for state attestation of graduates at all stages of education of the Republic of Uzbekistan.

It is essential that the standards be seen as a useful working document so that there is constant and reliable cross referencing between those working on syllabi, textbooks and assessment tools. Without this liaison, regular meetings and proper co-ordination between the working groups, the reform process in language teaching will not work. It will be necessary to provide support and proper training for the professionals involved in the design and delivery of these elements, and also, of course, to language teachers at all levels across the Republic. It is important to the success of the reform that these requirements are seen as aspirational rather than as a reflection of the status quo. Thus they may be seen as target levels, giving teachers and learners outcomes to aim for, even if the achievement of these targets takes a longer period of time.

REQUIREMENTS TEACHING AND LEARNING FOREIGN LANGUAGE

FOR LEVEL B2 Higher Education Non-Linguistic Undergraduate Programmes

Competences	CONTENT OF TEACHING AND LEARNING		
TOPIC BASED CONTENT	<ul style="list-style-type: none"> • Topics related to the Internet and ICT • Sociocultural topics on the specialism (specific comparisons and contrasts between Uzbekistan and target language countries, e.g. UK & USA, France, Germany). • Topics of specific/ professional purposes (background specialisation, trends in the specialisation) • Topics related to social life (social contact with the surrounding world). 		
	<ul style="list-style-type: none"> • Linguistic competence 	Competence in language skills	Listening
Speaking			<p>Spoken Interaction</p> <ul style="list-style-type: none"> • transactions • social talk and informal conversations • formal and informal discussions within and beyond the learner's specialism • chairing or leading a discussion • interviews • negotiations • telephone calls <p>Spoken Production (monologue).</p> <ul style="list-style-type: none"> • making reports • developing an argument, e.g. in discussion of a specialist topic • stating and supporting an opinion • making announcements • making a presentation on a specific topic • summarising an article, a discussion etc
Reading			<ul style="list-style-type: none"> • correspondence, including emails, notes and messages as well as letters • authentic texts containing specific material • texts containing specific lexis and terminology, e.g. abstracts, reports, extracts from textbooks • scientific and specialist literature (periodicals, E-literature) <p>Skills to be developed: reading for gist; reading for specific information; reading for detailed understanding; reading for orientation (signs, labels etc)</p>
Writing			<ul style="list-style-type: none"> • correspondence (letters, messages, etc) • specific reports (memos, CVs, etc.) • essays, summaries, abstracts, etc. • research papers (articles, final qualification works, etc.)

		Lexical competence	<ul style="list-style-type: none"> • development of specialist vocabulary and terminology, including common abbreviations • word fields at an intermediate/ upper-intermediate level • ways of word formation (compounds and affixation), international words and cognates/false cognates • antonyms, synonyms and other common lexical relationships
		Grammatical	largely accurate application of grammatical material covered at previous levels (verb tenses, modals, comparative degrees of adjectives and adverbs, determiners, prepositions etc) in general and academic contexts
Sociolinguistic competence			<ul style="list-style-type: none"> • intercultural awareness raising between Uzbek and other cultures (in both academic and social settings), and the way some of the issues are related to language, e.g. greetings, modes of address in academic and professional settings, basic politeness conventions in lectures, seminars etc. • further work on non-verbal elements of communication in different cultures: body language; non-verbal signals etc • email and messaging conventions in the foreign language as compared with L1
Pragmatic competence			<ul style="list-style-type: none"> • further development of presentation skills • linking ideas appropriately in spoken and written discourse • awareness of some of the degrees of formality of language needed in different social, academic and professional settings • strategies for interrupting, clarifying, paraphrasing, ‘repairing’ and compensating etc.

**REQUIREMENTS
FOR THE LEVEL OF GRADUATES ON FL
B2 LEVEL
Higher Education
Non-Linguistic Undergraduate Programmes**

Competences		Knowing/can do
LINGUISTIC COMPETENCE	COMPETENCE IN LANGUAGE SKILLS	<p>Listening</p> <p>Listen</p> <p><i>By the end of their undergraduate studies, learners can:</i></p> <ul style="list-style-type: none"> • understand and follow an extended talk or complex lines of arguments • can understand the essentials of lectures, talks and reports, detailed instructions and other forms of academic and professional presentations, questions and statements • understand announcements and messages • understand complex authentic speech in familiar and unfamiliar contexts • catch most of a conversation or discussion between target language speakers taking place around them • understand most radio, Internet and TV documentaries, interviews etc
		<p>Speaking</p> <p>Speak</p> <p>Spoken Interaction</p> <p><i>By the end of their undergraduate studies, learners can:</i></p> <ul style="list-style-type: none"> • negotiate with professional partners • make a request on specific area • engage in extended conversation with native speakers and sustain the interaction, taking the lead if necessary • take part in unprepared and natural discussions and debates • take part in an interview about their specialism • express their ideas and opinions clearly within the framework of a formal discussion • clarify, paraphrase and repair their own contributions to discussions

		<ul style="list-style-type: none"> • negotiate over a transaction or the resolution of a problem using appropriate levels of politeness and formality • ask and answer questions appropriately in formal settings, e.g. seminars <p>Spoken Production</p> <p>By the end of their undergraduate studies, learners can:</p> <ul style="list-style-type: none"> • give a well-structured presentation on a specific topic • give clear, detailed descriptions on a range of subjects related to their specialism • make an oral report on a specific topic • give a clear summary of an article, lecture or discussion • develop and sustain an argument on a familiar topic, supporting it with reasons, examples and evidence
	Reading	<p>By the end of their undergraduate studies, learners can:</p> <ul style="list-style-type: none"> • understand the main / specific points in information texts on familiar and unfamiliar topics • understand correspondence related to their interests or specialism • understand short descriptions of charts, graphs, tables • understand complex messages • understand and follow specific and complex written instructions or directions • locate specific information in longer articles and reports in their specialist field • read abstracts, conference programmes, contents pages etc, in order to decide whether to read certain sections or chapters for detail <p>Skills to be developed: reading for gist; reading for specific information; reading for detailed understanding; reading for orientation (signs, labels etc)</p>
	Writing	<p>By the end of their undergraduate studies, learners can:</p> <ul style="list-style-type: none"> • write specific messages (business letters, notes, e-mails) • write well-structured professional essays and reports, • write coherent scientific and research articles with a reasonable degree of accuracy and in an appropriate style (C1) • write proposals, summaries and abstracts • (if required) write final qualification works in their specialism (C1)
	Lexical Competence	<p>By the end of their undergraduate studies, learners can:</p> <ul style="list-style-type: none"> • use specific lexis and terminology in context • use topic-related vocabulary in communicative settings; • recognise and use a wide range of international words
	Grammatical Competence	<p>By the end of their undergraduate studies, learners can:</p> <ul style="list-style-type: none"> • use complex grammar and syntactical constructions in communicative settings • use appropriate linking words • analyse a piece of discourse in their own specialism to understand how it is structured in terms of cohesion and coherence

**CONTENT OF TEACHING AND LEARNING FOREIGN LANGUAGE
FOR LEVEL C1
Higher Education
Specialist Language Undergraduate Programmes**

Competences		CONTENT OF TEACHING AND LEARNING
TOPIC BASED CONTENT		<ul style="list-style-type: none"> • Topics related to everyday life (e.g. diet, bringing up children etc). • Topics related to wider society (e.g. ethical issues in medicine, science and technology, social issues, youth crime, community responsibilities etc.). • Topics of professional interest (e.g. Linguistic and Sociolinguistic issues, language teaching, language learning, etc.) • Sociocultural topics (acculturation, culture shock and social distance, preserving cultural identity, behaving in intercultural settings, being sensitive about culture while designing materials etc.)
Linguistic competence	Listening	<p>Learners should be exposed to:</p> <ul style="list-style-type: none"> • extended speech on abstract and complex topics; • a wide range of idiomatic expressions and colloquialisms and recorded and broadcast audio (e.g. in a station, sports stadium etc.); • complex interactions between third parties in group discussion; • television programmes and films • samples of authentic spoken language by both native speakers and non-native speakers, in formal, informal and academic settings • lectures, discussions and debates in their specialist field • classroom interaction in the target language
	Speaking	<p>Spoken Interaction (dialogue).</p> <ul style="list-style-type: none"> • inter-personal dialogues and conversations; • public debates and formal discussion; • lectures and talks on abstract and complex topics of a specialist nature beyond his/her own field; • job interview either as an interviewer or interviewee; • language in classroom settings <p>Spoken Production (monologue).</p> <ul style="list-style-type: none"> • instructions • presentations • developing and supporting an argument on concrete or abstract topics • expressing an opinion giving reasons • summarising an opinion, a discussion, a professional article etc
	Reading	<ul style="list-style-type: none"> • lengthy, complex texts likely to be encountered in social, professional or academic life; • books, fiction and non-fiction, including literary journals; • periodicals (magazines, newspapers); • instruction manuals (textbooks, cookbooks, etc.); • advertising material; • data including forms, teacher diary, questionnaires; • formal letters, emails etc • memoranda, reports, critical reviews and papers;

		Writing	<ul style="list-style-type: none"> • report articles • instructions for learning and teaching materials; • forms and questionnaires; • business and professional letters; • essays, reports, reviews; • qualification paper; • statement of intent; • CVs; covering letter.
		L exical competence	<ul style="list-style-type: none"> • broad lexical repertoire, idiomatic expressions and colloquialisms • contrasting and comparing specialist terminology (language teaching and applied linguistics) in Uzbek, Russian and the target language
		G rammatical	<ul style="list-style-type: none"> • complex sentences to convey meaning which is a central aspect of communicative competence; • grammatical semantics including grammatical elements, categories, structures and processes; • grammar at discourse level
Sociolinguistic Competence			<ul style="list-style-type: none"> • intercultural awareness raising between Uzbek and other cultures, and the way some of the issues are related to language, e.g. greetings, modes of address, basic politeness conventions etc. • further work on non-verbal elements of communication in different cultures: body language; non-verbal signals etc
Pragmatic Competence			<ul style="list-style-type: none"> • further development of presentation skills • linking ideas appropriately in spoken and written discourse • awareness of some of the degrees of formality of language needed in different social, academic and professional settings • strategies for interrupting, clarifying, paraphrasing, ‘repairing’ and compensating etc.

**REQUIREMENTS
FOR THE LEVEL OF GRADUATES ON FOREIGN LANGUAGE
C1 LEVEL
Higher Education
Specialist Language Undergraduate Programmes**

Competences		Graduates on FL CEFR C1
LINGUISTIC COMPETENCE	COMPETENCE IN LANGUAGE SKILLS	Listening
		<p><i>By the end of Year 4, graduates can:</i></p> <ul style="list-style-type: none"> • understand enough to follow extended speech on abstract and complex topics beyond his/her own field, though he/she may need to confirm occasional details, especially if the accent is unfamiliar; • recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts; • follow extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly; • easily follow complex interactions between third parties in group discussion and debate, even on abstract, complex unfamiliar topics; • understand a wide range of recorded and broadcast audio material, including some non-standard usage, and identify finer points of detail including implicit attitudes and relationships between speakers;

		<ul style="list-style-type: none"> • follow films employing a considerable degree of slang and idiomatic usage. • extract specific information from poor quality, audibly distorted public announcements, e.g. in a station, sports stadium etc.
	Speaking	<p>Spoken Interaction By the end of Year 4, graduates can:</p> <ul style="list-style-type: none"> • express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. • understand in detail speech on abstract and complex topics of a specialist nature beyond his/her own field, though he/she may need to confirm occasional details, especially if the accent is unfamiliar. • participate fully in an interview, as either interviewer or interviewee, expanding and developing the point being discussed fluently without any support, and handling interjections well; • easily keep up with and contribute to a debate, even on abstract, complex unfamiliar topics; • argue a formal position convincingly, responding to questions and comments and answering complex lines of counter argument fluently, spontaneously and appropriately <p>Spoken Production By the end of Year 4, graduates can:</p> <ul style="list-style-type: none"> • give elaborate descriptions and narratives by developing particular points and rounding off with an appropriate conclusion;
	Reading	<p>By the end of Year 4, graduates can:</p> <ul style="list-style-type: none"> • understand in detail lengthy, complex texts, whether or not they relate to his/her own area of specialism, provided he/she can reread difficult sections; • understand any correspondence given the occasional use of a dictionary; • understand in detail a wide range of lengthy, complex texts likely to be encountered in social, professional or academic life (e.g. professional articles, book chapters, reviews), identifying finer points of detail including attitudes and implied as well as stated opinions. • extract relevant detail from websites and journals related to their specialist field
	Writing	<p>By the end of Year 4, graduates can:</p> <ul style="list-style-type: none"> • write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion; • write clear, detailed, well-structured and developed descriptions and imaginative texts in an assured, personal, natural style appropriate to the reader in mind; • write clear, well-structured expositions of complex subjects, underlining the relevant salient issues (e.g. in a qualification paper); • expand and support points of view at some length with subsidiary points, reasons and relevant examples. • write personal text types (diaries, reflections etc) in an appropriate style • summarise in writing lectures, articles and discussions • write professional and academic reviews

LANGUAGE ABILITIES	Phonological Competence	<p><i>By the end of Year 4, graduates can:</i></p> <ul style="list-style-type: none"> • vary intonation and place sentence stress correctly in order to express finer shades of meaning.
	Lexical Competence	<p><i>By the end of Year 4, graduates can:</i></p> <ul style="list-style-type: none"> • have a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. • have a good command of idiomatic expressions and colloquialisms; • have occasional minor slips, but no significant vocabulary errors • have a good command and understanding of specialist language teaching and applied linguistics terminology in the target language
	Grammatical Competence	<p><i>By the end of Year 4, graduates can:</i></p> <ul style="list-style-type: none"> • consistently maintain a high degree of grammatical accuracy; errors are rare and difficult to spot

Appendix O: Example Syllabus for LAL for UZ EFL Teachers

Module	Vignette	Topic	Objectives
<p>Assessment-for-Learning</p> <p>Vignette – Personal Beliefs and Attitudes</p> <p>How your own beliefs/attitudes might influence ones' own assessment practices?</p>	<p>Shaholo: - Shaholo from Focus Group Four interjected the following into the group discussion when I asked the participants about how teachers' total score of continuous assessment were broken down: May I talk about my time when I was teaching in the time of the USSR? There was no continuous assessment. Because we had only two, three marks. [The marks were tests.] Either you came to class or you did not, and that was not [scored]. But nevertheless the level of the knowledge was much more back then. Now we have different criteria, sub-criteria, what the students should know during the lesson. We divide everything to the little, little details. Too much. All the items the students should acquire, let's say. But still, we pay a lot of attention. We all the time, we tried to modify this percentage, this points for continuous assessment. If you</p>	<p>Washback and Preparation</p>	<p>How assessments can influence teaching and learning materials</p> <p>How assessments can influence teaching and learning in the classroom</p> <p>How assessments can influence the design of a language course or curriculum</p>

	were not trained by British Council, who made PRESETT, you do not know exactly [how to score] [for continuous assessment] (Teacher 10_FG4).		
<p>Vignette – Personal Beliefs and Attitudes</p> <p>How your own knowledge of language assessment (for the classroom) might be further developed.</p>	<p>Aziza: More ESP teachers reported using presentations as part of a continuous assessment score because they believe that they are one of the most helpful techniques to learn specific language forms. Aziza, who is in her first year of teaching English at an engineering university in Karakalpakistan explained that she and her colleagues “... made students [do presentations]... although I can state it also, it was a bit difficult, but it was very helpful for their language” (Aziza_7). Aziza believed that speaking at the front of the room is an effective way for students to learn language. When I asked her what she meant by, “it was a bit difficult” (Aziza_7) Aziza emphasized that as a teacher, she found it hard to convince students that presentations are helpful for their language learning. Furthermore, I asked Aziza to clarify how she identifies her students’ language concerns during presentations, and how she then informs them afterwards of issues or concerns that they should continue to work on. She iterated that, “presentations help students practice speaking” (Aziza_7). I ended the conversation without pressuring her to be more specific about how she understood language skills to improve. Aziza’s interpretation of the use of presentations emphasizes a belief that <i>the more practice, the better for language development</i>, but, she seemed unable to clarify how she identified her students’</p>	Assessment in Language Pedagogy	<p>How to diagnose learners’ strengths and weaknesses</p> <p>How to give useful feedback on the basis of an assessment</p> <p>How to use peer-/self- assessments</p>

	errors and/or how she informed them to better assist in their language development. Aziza believes that presentations are an effective way to learn language, but, she was unable to articulate how she used presentations to help students do so.		
<p>Vignette – Personal Beliefs and Attitudes</p> <p>How your own beliefs/attitudes may conflict with those of other groups involved in assessment.</p>	<p>Bibidaba: I talked with Bibidana and Olmar from Focus Group 3. Bibidana had been teaching English for 33 years in Bukhara and Olmar for 29 years in Andijan. Bibidana told the group that she does not usually use presentations in her languages class because it “takes up too [much] time with no real benefit” (Teacher 1_FG3). The no real benefit Bibidana is referring to is language skills. She views presentations as a waste of time because she believes that language skills are not being developed, and thus, she has a negative attitude toward presentations. When she does use a presentation activity to count as part of the continuous assessment score, she (and Olmar) score very similarly</p>	<p>Assessment Principles and Interpretation</p>	<p>The concept of validity</p> <p>The concept of reliability</p> <p>How to use assessments to evaluate achievement in language learning</p> <p>How to interpret what a particular score says about an individual’s language ability.</p>
<p>Assessment-of-Learning [Technical Skills of Assessment]</p> <p>Vignette – Assessment Policy and Local Practices</p> <p>The assessment (language testing) traditions in</p>	<p>Ajva And Aira: Some participants reported that they do not feel comfortable designing their own language tests because they do not have the necessary specialist knowledge and skills. Aisara and Ajva, who did not have a course in language assessment as part of their coursework for their Masters degree in TEFL, from Focus Group 1, discuss how they design tests.</p> <p>Aisara: The head of our department [...] can say, “You four or five people...yes?” “[You all are]</p>	<p>Developing and Administering Language Assessments</p>	<p>Aligning tests to proficiency frameworks (e.g., The common European framework of reference [CERR], American Council on the Teaching of Foreign language)</p> <p>From specifications to developing tests.</p>

<p>your local context</p>	<p>responsible for [making a] test.” [Then], [we each] have responsibilities.</p> <p>Ajva: We can find it from internet or we can make it ourselves. Yes, very often from internets...</p> <p>Interviewer: Can you please explain what you mean by ‘find it from internet?’</p> <p>Aisara: So [...] we are not test developers. ... That's why we try to take some activities [of] reading and writing [...] from some books and internets. We are not going to develop it [ourselves] because we are not experts. We just take [from the internet]. Of course, sometimes, [the contents] [are] connected with the material which we just [taught]. They are not the same [material, but, ones that we adapt].</p> <p>This discussion among Aisara, Ajava, and me, illustrates that the department head has considerable control over who designs tests. The participants reported going onto the internet to find items, reading passages, and/or questions that would best connect with what was taught in the course materials, because they did not feel</p>		<p>Section Outcome – Take a test that you use for your school and align it with the CEFR and explain how it is aligned.</p>
---------------------------	--	--	---

	<p>comfortable writing their own assessment items. Aisara in Focus Group 1 stressed that she is not a specialist test developer and felt uncomfortable creating test items from scratch. At the end of the focus group interview she said that she would “hope to learn more about the science of tests” (Aisara_FG1). Many teachers revealed they are responsible for designing tests but feel that they have insufficient background knowledge and/or skills to serve as test developers.</p>						
<p>Vignette – Assessment Policy and Local Practices</p> <p>The relevant legal regulations in a local context. (More than one person.)</p>	<p>In general, the participating Uzbekistan EFL teachers reported they do not use rubrics to score assignments, but instead, focus on their <i>impressions</i> of students’ work. Their impressions come from comparing students completed assignments to others in the classes. Additionally, the students’ <i>speaking</i> ability plays a role in how scores are assigned. I asked Mohira, a novice teacher who has been teaching EFL for one year in Kashkadarya, about how EFL teachers at her school score assignments. An excerpt of our conversation is presented below:</p> <table border="1" data-bbox="435 1360 914 1875"> <tr> <td data-bbox="435 1360 618 1581">Interviewer:</td> <td data-bbox="626 1360 914 1581">... what are some factors that influence a teacher's impressions on grading assignments?</td> </tr> <tr> <td data-bbox="435 1591 618 1875">Mohira:</td> <td data-bbox="626 1591 914 1875">Mostly they refer to speaking abilities. It doesn't matter if even they're teaching writing. For example, if I am teaching writing,</td> </tr> </table>	Interviewer:	... what are some factors that influence a teacher's impressions on grading assignments?	Mohira:	Mostly they refer to speaking abilities. It doesn't matter if even they're teaching writing. For example, if I am teaching writing,	Scoring and rating	<p>Scoring closed-response questions (e.g., Multiple Choice Items)</p> <p>Using Rating Scales to Score Speaking or Writing Performance</p> <p>Scoring open-ended questions (e.g., short answer questions)</p>
Interviewer:	... what are some factors that influence a teacher's impressions on grading assignments?						
Mohira:	Mostly they refer to speaking abilities. It doesn't matter if even they're teaching writing. For example, if I am teaching writing,						

		<p>there are some students who have been in UK, in US, and they come, but they don't have appropriate writing skills. They have spelling mistakes. We teachers don't have [appropriate] writing skills... So, I will provide a [score] with their speaking.</p>		
<p>Vignette – Assessment Policy and Local Practices</p> <p>How to determine if the results from a language assessment are relevant to the local context.</p>	<p>Below is a conversation I had with Ulugbek from Djizzak:</p>	<p>Ulugbek: After tests are given [to students to take], I have seen many tests that [I believe] are poorly written, many times, and I say something.</p> <p>Interviewer: Tell me more about that?</p> <p>Ulugbek: I have seen mistakes in the test [papers], which were not edited properly before the exam. I have seen [problems] in those [test papers which were created by other teachers]. [If a problem in the test paper was recognized] we would not lower the score of a student, because, it was our mistake, [the] teacher's mistake, and we [will] give</p>	<p>Statistical Research</p>	<p>Using statistics to analyze overall scores on a particular assessment</p> <p>Using statistics to analyze the difficulty of individual items (questions)/tasks</p> <p>Using statistics to analyze the quality of individual items</p> <p>Section Outcome: Analyze a data set about Objective and subjective Sections.</p>

	<p>[the student] one extra point [...], so we add scores for this particular question which contains a mistake, so I have seen these kinds of problems.</p> <p>Interviewer: When you bring up these problems, do you as a group, I guess with the people that make the exam, talk about how to fix it for next time?</p> <p>Ulugbek: Mm-hmm (affirmative). Yes, we do. I think we should do some math [conduct some statistical analyses of some test items], but no one really knows [no one is available who knows how to do this.]</p>		
--	---	--	--