

## University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]





# **University of Southampton**

FACULTY OF ARTS AND HUMANITIES

MODERN LANGUAGES AND LINGUISTICS

**LANGUAGE ASSESSMENT LITERACY DEVELOPMENT:  
EXPLORING, EXPERIMENTING AND REFLECTING ON THE STATUS QUO OF EFL  
TEACHERS AMONG EDUCATIONAL SECTORS IN TAMAULIPAS, MEXICO**

BY

**RICARDO DE LA GARZA CANO**

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

MAY 2020



## **Abstract**

Faculty of Arts and Humanities

Modern Languages and Linguistics

Thesis for the degree of Doctor of Philosophy

Language Assessment Literacy Development:

Exploring, Experimenting and Reflecting on the Status quo of EFL Teachers among Educational Sectors in Tamaulipas, Mexico

by

Ricardo de la Garza Cano

There have been significant recommendations from researchers and international organisations to improve teachers' assessment skills and practices in educational institutions with the intention to maximise and provide satisfactory opportunities for instruction and learning in the classrooms. (AFT, 1990; Stiggins, 1991; Herrera & Macias, 2015). Language Assessment Literacy (LAL) refers to the foundations of the uses and applications of appropriate knowledge, technical considerations and school-based assessment practices concerning the use of language tests and other alternative assessments (Stiggins, 2002; Scarino, 2013; O'Sullivan, 2017). Research on LAL thus far has reported a lack of assessment training among language teachers in developing, administering and interpreting test-takers' scores across the globe (Quilter & Gallini, 2000; Vogt & Tsagari, 2014). In addition, some academic programmes such as Applied Linguistics, TESOL and professional development courses may not include sufficient contents for teachers to become well-skilled in the field of language testing and assessment (Davies, 2008; Lam, 2015).

A significant number of studies that have been conducted to explore teachers' language assessment literacy have investigated different areas like knowledge of assessment, technical aspects and classroom assessment practices. However, little attention has been given to their teaching contexts and how different educational sectors operate assessments in their institutions. There are some factors that might influence teachers' language assessment literacy. These are their previous assessment experiences, academic background, practicum and developmental trajectories, as well as the institutional operational assessment system in their institutions. Professional development trainings have largely assisted teachers with teaching-oriented opportunities to bring appropriate knowledge, skills and practices to the classroom. Nevertheless, a research gap has been identified to explore the relationship between EFL teachers and language

program managers among their educational sectors (e.g., the public, private and language institutes) towards the field of language assessment literacy.

This dual-purpose research study aims at: a) examining the status quo of language assessment literacy in EFL teachers and language programme managers in Tamaulipas, Mexico with the intention to explore how teachers self-perceived their language assessment skills and practices (TLASP) in relation to their educational sectors, as well as b) measuring the EFL teachers' progress and impact from a workshop to improve their LAL. Two types of research methodologies were applied: exploratory and experimental. A mixed-method research study was carried out. The data collection procedures were through a survey (n=315) to examine participants backgrounds. Then, semi-structured interviews were conducted with EFL teachers and language programme managers (n=15). Also, a contextualised workshop was developed based on the survey and interviews outcomes and this workshop was administered to participants (n=10). Lastly, a focus group interview (n=6) was conducted to a) examine the qualities of the workshop in terms of the quality of the workshop and the areas where participants became aware of knowledge and practices and b) to measure any direct impact of the workshop towards participants' institutional contexts and professional development.

Findings showed that EFL teachers in Tamaulipas differed from their self-perceived assessment repertoire among the three educational sectors and it was within their institutions where teachers could expand their assessment literacy levels to explore and gain more knowledge and skills, as well as fundamental practices in language testing and assessment. This investigation has clear pedagogical and institutional implications on how EFL teachers use assessments and tests to maximise instruction and learning. In addition, this research study aimed to meet the future research calls from researchers in relation to explore and expand the areas of what teachers should know and do with assessments as individuals (teachers), classrooms, institutions and educational sectors at a State Level.

# Table of Contents

<b>Table of Contents</b> .....	<b>i</b>
<b>Table of Tables</b> .....	<b>vii</b>
<b>Table of Figures</b> .....	<b>ix</b>
<b>Research Thesis: Declaration of Authorship</b> .....	<b>xi</b>
<b>Acknowledgements</b> .....	<b>xiii</b>
<b>Abbreviations</b> .....	<b>xv</b>
<b>Chapter 1 Introduction</b> .....	<b>1</b>
1.1 Personal Motivation .....	1
<b>1.2 Background of the Study</b> .....	<b>2</b>
1.3 Rationale of the Study .....	4
1.4 Purpose of the Study .....	6
1.4.1 Research Questions.....	8
1.5 Scope of the Study.....	9
1.6 Significance of the Study .....	10
1.7 Structure of the Thesis .....	11
<b>Chapter 2 Language Assessment Literacy</b> .....	<b>13</b>
2.1 Literacy and Assessment .....	13
2.2 Language Assessment Literacy.....	17
2.3 Language Assessment Literacy: Stakeholders and Dimensions.....	22
2.4 Teachers' Language Assessment Skills and Practices (TLASP) .....	24
2.4.1 Knowledge of Assessment.....	24
2.4.1.1 Applied Linguistics .....	25
2.4.1.2 Educational Assessment.....	26
2.4.1.3 Institutional Assessment Policies .....	32
2.4.2 Technical Considerations in Assessments .....	34
2.4.2.1 Testing Principles.....	35
2.4.2.2 Assessment Purposes .....	38
2.4.2.3 Assessment Processes .....	40

## Table of Contents

2.4.2.4	Educational Measurement.....	43
2.4.3	School-Based Language Assessments.....	45
2.4.3.1	Language Assessment .....	46
2.4.3.2	Assessment for Learning.....	48
2.4.3.3	Communication of Assessments and Results .....	50
2.4.3.4	Educational Support through Assessment.....	51
2.5	Contextualising Educational Assessment: Mexico and Tamaulipas .....	53
2.5.1	Chronosystem .....	54
2.5.2	Macrosystem .....	56
2.5.3	Exosystem .....	58
2.5.4	Microsystem .....	60
2.5.5	Institutional (The School).....	62
2.5.6	Individual (The Teacher) .....	65
2.5.6.1	Previous Assessment Experiences .....	65
2.5.6.2	Teachers' Professional and Academic Trajectories .....	68
2.6	Review of Empirical Studies on Language Assessment Literacy.....	72
2.7	Summary .....	78
<b>Chapter 3</b>	<b>Methodology .....</b>	<b>81</b>
3.1	Research Methodology.....	81
3.2	State of Tamaulipas: Institutions and Educational Sectors .....	84
3.2.1	Educational Institutions .....	86
3.2.2	Educational Sectors.....	86
<b>3.3</b>	<b>Participants' Profiles and Samples .....</b>	<b>87</b>
3.3.1	First Phase (Survey).....	88
<b>3.3.2</b>	<b>Second Phase (Interviews) .....</b>	<b>90</b>
<b>3.3.3</b>	<b>Third Phase (Instruction of the Workshop) .....</b>	<b>91</b>
<b>3.3.4</b>	<b>Fourth Phase (Focus Group Interviews) .....</b>	<b>92</b>
3.4	Research Ethical Consideration .....	93
3.5	Development of the Research Instruments and Interviews.....	93
3.5.1	Teachers' Language Assessment Literacy Development Inventory (TALDI)....	94

3.5.1.1	Survey Development .....	94
3.5.1.2	Pilot Study 1.....	96
3.5.1.3	Final Version of TALDI.....	98
3.5.2	Semi-structured Interviews.....	102
3.5.3	Instruction of the Workshop in Language Testing and Assessment.....	104
3.5.3.1	Pre-workshop and Pilot Study 2 .....	105
3.5.3.2	During the Workshop .....	109
3.5.3.3	Syllabus.....	109
3.5.4	Focus Group Interview .....	113
3.6	Data Analysis Methods.....	114
3.6.1	Phase 1 (Survey: TALDI).....	114
3.6.2	Phase 2 (Interviews) .....	116
3.6.3	Phase 3 (Workshop) .....	117
<b>3.6.4</b>	<b>Phase 4 (Focus Group) .....</b>	<b>117</b>
3.7	Summary of the Chapter .....	118
<b>Chapter 4</b>	<b>Survey (TALDI) Findings.....</b>	<b>121</b>
4.1	Participants' Educational Backgrounds.....	121
4.2	Participants' Assessment Experiences and Educational Sectors .....	123
4.2.1	Test Development and Responsibilities .....	124
4.2.2	Language Skills in Assessments.....	125
4.2.3	Alternative Assessments .....	125
4.2.4	Ethical Considerations in Language Assessments.....	126
4.2.5	Teachers' Previous and Current Experiences in Assessment.....	127
4.2.6	Critical Assessments Problems in Institutions.....	127
4.3	Second Language Testing and Assessment Backgrounds .....	128
4.4	Participants' Continuing Professional Development .....	130
4.5	Teachers' Assessment Skills and Practices.....	133
4.6	Summary of the Chapter .....	139
<b>Chapter 5</b>	<b>Semi-structured Interview Findings .....</b>	<b>143</b>
5.1	Language Tests (Relevance and Occurrences).....	143

## Table of Contents

5.2	Test Development: Language Skills and Linguistic Components.....	145
<b>5.3</b>	<b>Continuing Professional Development and Language Assessment .....</b>	<b>148</b>
<b>5.3.1</b>	<b>Professional Development: Trainings and Qualifications .....</b>	<b>148</b>
<b>5.3.2</b>	<b>Professional Development in Language Assessment Literacy .....</b>	<b>150</b>
<b>5.3.3</b>	<b>Lack of Professional Development in Language Assessment Literacy .....</b>	<b>151</b>
5.4	Teachers' Previous and Current Assessment Experiences .....	154
5.5	Teachers' Assessment Knowledge and Technical Considerations.....	156
5.6	Teachers' School-based Assessments.....	161
<b>5.7</b>	<b>Critical Issues among Educational Sectors .....</b>	<b>163</b>
<b>5.7.1</b>	<b>The Public Sector .....</b>	<b>163</b>
<b>5.7.2</b>	<b>The Private Sector .....</b>	<b>164</b>
<b>5.7.3</b>	<b>Language Institutes .....</b>	<b>165</b>
<b>5.8</b>	<b>Summary of the Chapter.....</b>	<b>165</b>
<b>Chapter 6</b>	<b>Workshop: Progress and Impact.....</b>	<b>169</b>
<b>6.1</b>	<b>Pre- and Post- Measure.....</b>	<b>169</b>
6.2	Reflective and Practical Tasks .....	171
<b>6.2.1</b>	<b>The Public Sector .....</b>	<b>172</b>
<b>6.2.2</b>	<b>The Private Sector .....</b>	<b>175</b>
<b>6.2.3</b>	<b>The Language Institute Sector .....</b>	<b>177</b>
<b>6.3</b>	<b>Workshop Evaluation Sheet and Focus Group Interview .....</b>	<b>179</b>
6.4	Impact of the Workshop.....	185
<b>6.4.1</b>	<b>Participants' Practicum Trajectories.....</b>	<b>185</b>
<b>6.4.2</b>	<b>Participants' Developmental Trajectories.....</b>	<b>187</b>
<b>6.4.3</b>	<b>Participants' Institutional Operational Assessment Systems.....</b>	<b>188</b>
<b>6.4.4</b>	<b>Participants' Local and National Considerations .....</b>	<b>189</b>
6.5	Summary of the Chapter.....	190
<b>Chapter 7</b>	<b>Discussions and Conclusions .....</b>	<b>193</b>
7.1	Status Quo of Teachers' Language Assessment Literacy in Tamaulipas .....	193
7.1.1	Test Relevance and Occurrences .....	194
7.1.2	Test Development and Responsibilities.....	195

7.1.3	Assessment of Language Skills and Linguistic Features .....	196
7.2	Self-perceive Teachers' Language Assessment Skills and Practices (TLASP) .....	198
7.2.1	Knowledge of Assessment.....	198
7.2.2	Technical Considerations in Assessment .....	200
7.2.3	School-Based Language Assessments .....	201
7.3	Professional Development and Language Assessment Backgrounds.....	204
7.4	Workshop: Progress and Areas of Improvement.....	207
7.5	Impact of the Workshop .....	208
7.6	Contributions from these Research Studies .....	211
7.6.1	Contribution towards Language Assessment Literacy.....	211
7.6.2	Contribution towards Professional Development .....	212
7.6.3	Contributions towards Methodological Decisions.....	215
7.7	Limitations of the Study .....	215
7.8	Opportunities for Future Research .....	216
	<b>Appendix A Participants' Consent Form .....</b>	<b>219</b>
	<b>Appendix B Participant Information Sheet .....</b>	<b>220</b>
	<b>Appendix C Example of an Approval Participation.....</b>	<b>222</b>
	<b>Appendix D Survey (TALDI) .....</b>	<b>225</b>
	<b>Appendix E Interview Guidelines .....</b>	<b>232</b>
	<b>Appendix F Example of an Interviewee (LI-T4) .....</b>	<b>233</b>
	<b>Appendix G Syllabus of the Workshop.....</b>	<b>238</b>
	<b>Appendix H Pre- and Post- Measure.....</b>	<b>243</b>
	<b>Appendix I Example of Practical and Reflective Tasks.....</b>	<b>248</b>
	<b>Appendix J Workshop Evaluation Sheet .....</b>	<b>249</b>
	<b>Appendix K Focus Group Interview Guidelines .....</b>	<b>251</b>
	<b>Appendix L Findings from Participants' Educational Sectors.....</b>	<b>252</b>
	<b>Appendix M Participants' Individual Scores from Pre- and Post- Measure .....</b>	<b>253</b>
	<b>Appendix N Language Assessment Literacy Development Framework (Descriptors)....</b>	<b>255</b>
	<b>List of References .....</b>	<b>261</b>



## Table of Tables

<b>Table 1 Berry &amp; O’Sullivan’s (2016) Stakeholders Classification</b> .....	23
<b>Table 2 Levels of Assessment Literacy</b> .....	23
<b>Table 3 Matrix of Knowledge of Assessment</b> .....	34
<b>Table 4 Testing Principles: Definitions and Applications</b> .....	35
<b>Table 5 Purposes of Classroom-based Assessment and Applications (Popham, 2011)</b> .....	39
<b>Table 6 Matrix for Technical Consideration in Assessments</b> .....	45
<b>Table 7 Types of Responses and their Applications (Brown, 2012: 135)</b> .....	47
<b>Table 8 Performance Assessments vs Standardised Tests (Hamp-Lyons, 1997)</b> .....	49
<b>Table 9 Other Assessment Tools and Their Applications</b> .....	50
<b>Table 10 Phases of Communication of Assessments (Brown, 2012)</b> .....	51
<b>Table 11 Matrix of School-based Assessments Component</b> .....	52
<b>Table 12 Continuing Professional Development Framework (British Council, 2015)</b> .....	71
<b>Table 13 Research Design, Data Collection Phases and Chronological Description</b> .....	84
<b>Table 14 Participants from Phase 1 Survey: (TALDI n=315)</b> .....	89
<b>Table 15 Participants from Phase 2 (Interviews)</b> .....	90
<b>Table 16 Participants from Phase 3 (Workshop)</b> .....	92
<b>Table 17 Participants from Phase 4 (Focus Group)</b> .....	92
<b>Table 18 Teachers’ Assessment Skills and Practices</b> .....	102
<b>Table 19 Findings for the Selection of Contents for the Workshop (N=315)</b> .....	107
<b>Table 20 Contents and Items of the Measure</b> .....	110
<b>Table 21 Reflective Tasks and Workshop Sessions</b> .....	111
<b>Table 22 Practical Tasks and Workshop Sessions</b> .....	111

Table of Tables

<b>Table 23 Triangulation of the Mixed-method Approaches</b> .....	116
<b>Table 24 Research Study Overview</b> .....	120
<b>Table 25 Participants' Academic Profiles</b> .....	122
<b>Table 26 Test Responsibilities among Educational Sectors (n=268)</b> .....	124
<b>Table 27 Matrix of Language Skills within Tests (n=256)</b> .....	125
<b>Table 28 Overall Participants' Needs for the Workshop</b> .....	130
<b>Table 29 Overall Findings of Professional Development Features</b> .....	132
<b>Table 30 Overall Findings from Participants Assessment Skills and Practices</b> .....	134
<b>Table 31 Multiple Comparisons among Participants' Educational Sectors</b> .....	136
<b>Table 32 Overall Findings from Pre- and Post- Measures (n=10)</b> .....	170
<b>Table 33 Findings from Participants Educational Sector (n=10)</b> .....	170
<b>Table 34 Workshop Evaluation Sheet (WES) Analysis</b> .....	180

## Table of Figures

<b>Figure 1 Brown and Abeywickrama's (2010:6) Diagram</b> .....	15
<b>Figure 2 Language Assessment Literacy Standards for the Language Classroom</b> .....	24
<b>Figure 3 Assessment Pyramid</b> .....	30
<b>Figure 4 Fulcher and Davidson's (2010) Test Development Framework</b> .....	41
<b>Figure 5 Multilayer Analysis of Mexican Assessment Policies in Education</b> .....	54
<b>Figure 6 Development of the Workshop.</b> .....	105
<b>Figure 7 Data Analysis of Phase 3 and 4</b> .....	118
<b>Figure 8 Teachers' Language Assessment Literacy Development Framework (TLALDF)</b> .....	213



## Research Thesis: Declaration of Authorship

Print name: Ricardo de la Garza Cano

Title of thesis: Language Assessment Literacy Development: Exploring, Experimenting and Reflecting on the Status quo of EFL Teachers among Educational Sectors in Tamaulipas, Mexico.

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signature: ..... Date: 05/07/2021



## Acknowledgements

First of all, I would like to thank Dr Ying Zheng for her encouragement during these six years while I was studying at the University of Southampton. Her guidance and feedback inspired me to pursue a PhD under her supervision with a great deal of enthusiasm. Although our journey has had ups and downs, Dr Zheng has always helped me overcome my own fears inside and outside academia. Her constant support has opened a large number of doors in conferences in which I have had the opportunity of meeting the most significant and relevant researchers in assessment across Europe and China.

This achievement would not have been possible without my sponsors. The National Council for Science and Technology (CONACYT) and the Tamaulipas Council for Science and Technology (COTACYT). Besides this, The National and Local Ministries of Education (SEP and SET respectively). Similarly, I would like to thank to the Autonomous University of Tamaulipas (UAT) and the Centre for Languages and Applied Linguistics (CELLAP) Campus Victoria for their substantial support during this journey studying overseas.

Conducting this research and writing this thesis have been, by far, the most complex and challenging process I have had thus experienced. These past years have been difficult, not solely because of the complexity of this project, but for a significant number of unexpected situations related to personal issues that were presented along this journey. However, I had been blessed to have my parents safe and sound, Emilio M. De la Garza Gonzalez and Silvia Cano Cano, who have always inspired me to be a good man and to never give up regardless the unforeseen risks, challenges and personal storms. Also, I would like to thank my sister, Silvia de la Garza Cano, who is not only the eldest sister, but in fact, a friend who was there in the darkest times of this journey. To my brother, Emilio de la Garza Cano, who has been an example to forge this thesis and see the progress of an educator who has a different educational and professional trajectories to become an excellent accountancy teacher. Their support has been a vital element of persistence, although they were miles away in Mexico.

Coming to this country was not just an academic opportunity, in fact, it widely opened my eyes to value things like education, friendship and love for others. As a result, I would like to thank Alvaro Cortes, Dr Amina Doudi, Dr Aneesha Seti, Barry Shaw, Dr Daniel Cudeiro (+), Dimitra Banti, José (Pepe) Rodriguez and Antoni (Toni) Martinez-Brown for making this adventure unique. In addition, to my Mexican friends who always heartened me and gave me their positive vibes to be where I am, to Aida Rodriguez, Dr Fernanda Gonzalez, Mitzy Valdez de Méndez and Roberto Méndez.



## Abbreviations

**CAE:** Certificate of Advance English

**CELTA:** Certificate of English for Language Teaching to Adults

**CPD:** Continuing Professional Development

**FCE:** First Certificate of English

**FG:** Focus Group

**ICELT:** In-service Certificate of English Language Teaching

**IELTS:** International English Language Testing System

**INEE:** National Institute for Educational Evaluation

**INT:** Interview(s)

**KET:** Key English Test

**LI:** Language Institute

**LTA:** Language Testing and Assessment

**MEXTESOL:** Mexican Association for Teachers of English to Speakers of Other Languages

**PET:** Preliminary English Test

**PRIV:** The private sector

**PUB:** The public sector

**REF:** Reflections

**SEP:** Secretaría de Educación Pública (Ministry of Education in Mexico)

**SET:** Secretaría de Educación de Tamaulipas (Ministry of Education in Tamaulipas)

**TESOL:** Teachers of English to Speakers of Other Languages

**TOEFL:** Test of English as a Foreign Language

**TKT:** Teaching Knowledge Test

**TLASP:** Teachers' Language Assessment Skills and Practices

**WES:** Workshop Evaluation Sheet



# Chapter 1 Introduction

Tests, assessments and evaluations are considered fundamental features which could influence our learning, performances and professional trajectories. Although tests and assessments are mostly seen and used in our educational journeys, these tools could have a significant impact outside the classrooms (McNamara, 2000). As human beings, we occasionally evaluate and reflect on different decisions that impact our daily life; for example, deciding which educational programme meets our personal needs, seeking better job opportunities or to demonstrate enough linguistic competence to migrate from one country to another (Fulcher, 2012).

This Chapter introduces the research project by presenting my own personal motivation regarding to the fields of educational assessment and language proficiency. Then, Section 1.2 highlights the background of this study and defines the concept of language assessment literacy. Section 1.3 presents the rationale of the study by detailing the current educational assessments practices in Mexico and presents the research gaps. The purpose of the study is discussed in Section 1.4 with the research questions which were addressed based on the research aims. Section 1.5 denotes the scope of the study and the theoretical background of language assessment literacy and professional development. This sub-section also discusses the methodological approaches on how and where this study was conducted. Section 1.6 outlines the significance of this research conducted in the Mexican State of Tamaulipas and the possible contributions towards the study of language assessment literacy and professional development, as well as, the practical contributions towards the language educational system in Mexico. Lastly, this chapter concludes with a comprehensive summary of the structure of this thesis labelling each chapter's contents.

## 1.1 Personal Motivation

My motivation behind this research on educational assessment started when I was just a child in a public primary school at the age of 6, where I encountered my first classroom tests. During these six years of my primary education, I was required to pass approximately 48 basic modules (e.g., Spanish, Mathematics, History, Geography, English as a Foreign Language, etc....) and took at least 440 tests and one National Measure (standardise test) for the admission to the secondary education. I considered myself as a dedicated pupil who aimed to achieve high grades to be in the top tier of my class, however, I was unable to meet these standards due to unfairness. My mother, who used to work in another primary school under the public sector, would give me numerous other learning materials to improve my understanding in different topics in the hope to

## Chapter 1

achieve better scores, but I was still unable to improve on my performance in those tests. Years passed and I progressed from primary school to start secondary school. Notoriously, those classmates, who continuously got higher marks in primary school, were achieving lower marks than they were expecting, whereas I, on the other hand, was starting to receive higher marks during these three years as I was performing better in secondary school assessments than primary and was part of the top tier for three years. My performance in further education mirrored that of secondary school, obtaining consistently high marks throughout my time there. Based on this experience I noticed that sometimes teachers do not prepare their assessments well, and this could bring a lack of opportunities to empower different skills. During my upper-secondary school, I started studying the English language in a language institute. It was in here where my motivation towards this language increase since I was taking classes with well-prepared teachers. My enthusiasm towards the English language increased during the upper-secondary school that ultimately enabled me to study a Bachelor's degree related to English language teaching in which I was awarded first place during my tertiary studies.

I studied a bilingual programme in Applied Linguistics in the Mexican state of Tamaulipas. The inspiration about educational assessment grew towards how language teachers could measure someone's language ability and performances. As a language learner and teacher, I was intrigued by how teachers used to mark their students, some of them were very thorough providing relevant and constructive feedback, whereas others just corrected students' mistakes, providing very little feedback aside. During my Master's degree programme at the University of Southampton, the module of language proficiency assessment became the most remarkable and stimulating course. I realised that studying in another country, the national standards of assessments were different and since I was committed to study there for a while, I had to comprehend the assessment practices that my lecturers and professors used towards my own assessments. As a result, upon successful completion of my Masters, I had the opportunity to start a Philosophy Doctorate in Second Language Testing and Assessment, specifically relating to assessment. It was in there where I decided to examine and explore how much English as a Foreign Language (EFL) teachers know and do about assessment skills and practices. Since I was born in the State of Tamaulipas, I decided to scrutinise teachers' assessment skills and practices among different educational contexts.

### **1.2 Background of the Study**

Assessment literacy is a concept introduced by Richard Stiggins (1991), which states that educators and teachers need to take into consideration several assessment skills, practices and, most importantly, knowledge to administer appropriate assessments in their institutions (Stiggins,

1999; Inbar-Lourie, 2008a; O' Loughlin, 2013). Stiggins also highlighted that teachers with a suitable assessment literacy level could straightforwardly distinguish among their assessment practices: what they assess, how they assess and why they assess. From this point of view, educators, teacher trainers and programme managers could promptly identify and prevent inaccuracies in assessments and other educational practices (Stiggins, 1995). Similarly, assessment literacy also focuses on a variety of methods to collect students' learning evidence and strategic approaches for delivering constructive feedback (Inbar-Lourie, 2008b; Fulcher, 2012; Taylor, 2013). In other words, this concept is considered as a repertoire of knowledge of assessment competencies, techniques, experiences towards a better use of tests and assessments (Boyles, 2005; Hoyt, 2005; Malone, 2013).

Therefore, with new emerging local and international reforms in education, there has been a global demand to grow awareness among teachers to improve their assessment knowledge, techniques and practices to increase their educational standards (Stiggins, 2005; OECD, 2010; Malone 2013). International organisations, researchers and several educational institutions have been investigating different features of language assessment towards better learning, teaching and assessing methods for language instruction. Designing instruments with the intention of exploring and analysing current teachers' assessment literacy has been a complex process due to the difficult process to distinguish what teachers should know about assessment. It is important to understand the context and the purpose of these instruments, but more notably, it is crucial to adapt several assessment principles or standards accordingly to the stakeholders' current needs. An abundance of instruments to analyse whether teachers are assessment literates had previously been developed and adapted to observe how these practices are dealt with among different stakeholders (e.g., students, teachers, parents, etc.).

The first attempt to highlight what teachers should know about testing and assessments was based on the American Federation of Teachers (AFT) in 1990 that outlined seven vital standards: 1) select assessment methods, 2) develop assessments for the classroom, 3) administer and scoring tests, 4) use scores to aid instructional decisions, 5) communicate results, 6) be aware of inappropriate and 7) unethical uses of tests (AFT, 1990). Numerous studies have used these assessment literacy standards as reference for investigating assessment features, such as: teachers' assessment knowledge (e.g., Popham, 1991; Plake & Impara, 1993; Inbar-Lourie, 2008a, 2013), assessment training and professional development (e.g., Jeong & Ballera, 2013, Popham, 2006), and test development (e.g., Rodriguez & Zabala, 2016).

Almost 30 years later from the first attempt to define what language assessment literacy was, DeLuca, LaPointe-McEwan and Luhanga (2016) conducted one of the most significant

## Chapter 1

document analyses, from my point of view, in relation to the core elements for assessment literacy. They analysed a significant number of assessment standards across the globe. They investigated 15 educational assessment standards from developed countries like the United Kingdom, Australia, the United States of America and New Zealand. These researchers categorised a total of eight teachers' language assessment literacy standards, including one more since the American Federation of Teachers: 1) assessment purposes, 2) assessment processes, 3) communication of assessment results, 4) assessment fairness, 5) assessment ethics, 6) measurement theory, 7) assessment for learning, and 8) education and support for teachers. The purpose of this extensive research was to identify the psychometric properties of 36 assessment literacy instruments and develop a framework which these authors prioritise and include areas such as assessment for learning and educational support through assessments. Some of the relevant outcomes of this document analysis was to highlight three main stages of evolution of the standards from a perspective of high-stakes tests towards classroom assessments and to enforce teachers' current needs to empower their teaching and learning methods. Nevertheless, there is still a lack of understanding what teachers should know and do regarding language assessment in their current educational settings and classrooms.

### **1.3 Rationale of the Study**

Classroom teachers are expected to use a variety of techniques, skills and appropriate knowledge for developing testing instruments and to administer assessment practices that enable a constant improvement of learning and teaching practices. Language testing and school-based assessments, those tests and other methods to assess students' proficiency, are the two dominant aspects of language education and the main agents are teachers and learners, who are primarily involved in this educational practice (Taylor, 2009; Coombe et al., 2012). It has been suggested that assessments provides clear evidence to make critical judgments to improve teachers' assessment practices in the classroom and to increase the institutional profile with the assistance and support from their institutional authorities (Taylor, 2009).

In the field of applied linguistics, appropriate assessment knowledge, skills and practices are important elements for any language teacher regardless of their educational programme. Language Assessment Literacy refers to the specific knowledge, practices and skills that gives language teachers the prerequisites to identify and prevent inaccuracies in relation to the development, administration and interpretation of language tests (Fulcher, 2012; Vogt and Tsagari, 2014). Teachers' assessment knowledge, technical considerations and classroom practices are essential to distinguish between appropriate and inappropriate use of language assessments (Davison, 2004). Teachers need to be aware of strategies and methods to track

students' progress to identify whether the teaching methods have been suitable for their students. Research studies have also demonstrated that by developing relevant and appropriate testing resources, language teachers could become aware in which areas of teaching, learning and assessing language require to amend with the clear intention to maximise students' performances (Thomas, Allman & Beech, 2004; Herrera & Macias, 2015).

Language teachers have wide responsibilities inside the classroom. One of these is to deliver knowledge to students through different teaching strategies and methods to meet the learning objectives of their courses. Arguably, the most essential is the adequate assessments of their students' performances and their progression trajectories to identify whether the objectives of a course has been met (Thomas et al., 2004; Khadijeh & Amir, 2015; Volante & Fazio, 2007). It has been indicated that the most prominent feature that impacts on students' learning pathways and their progression is the *'teacher'* (William, 2006; Soneville, 2007; Mertler, 2009). As a consequence, teachers play a crucial role in language assessment practices, even though that teachers spend a large amount of time conducting assessment practices in the classroom, these assessments might not be adequate to their students. Spolsky stated that *'if an educator is good in teaching, he or she is good in assessing the learners as well'* (Spolsky, 1978, cited in Khadijeh & Amir, 2015:2). The previous quotation highlights that educational assessment is an influential factor to define the qualities of a good teacher, but that does not always mean that a good teacher might be a good assessor as well. Essential knowledge and constant practice on how to teach and how to assess the language are two different fields in which teachers should continuously evaluate their competences to empower and highlight their educational success. Assessment, as has been stated before, is also included in students' learning opportunities. Eckhout, et al., (2005) stated, *'Good teaching is impossible in the absence of good assessment'* (2005:4). Therefore, we can assume that teaching and assessing the language are two parallel stages, but they need to be complemented each other.

Numerous studies have examined and analysed different features of language assessment literacy in various educational contexts. Researchers have suggested that instructors should develop essential knowledge of assessment to improve their teaching competencies and to expand their practices towards their language education. Nevertheless, literature suggests that a large proportion of language teachers are not well-prepared with these vital elements to assess students accordingly to the individual and programme needs (Taylor, 2009; Davison & Leung, 2009; Davison & Hamp-Lyons, 2010). From this point, researchers and organisations across the globe have explored which essentials features are needed for educators, as well as for other relevant stakeholders, to know about language testing and classroom-based assessment practices.

## 1.4 Purpose of the Study

Educational assessment has caught researchers', teachers' and practitioners' attention to reinforce and improve the use of tests, not only inside a classroom, but also to examine and observe particular features at institutional, local and national scales. It is important to mention that significant attention has been given to *high-stake* tests or “*standardised tests*” worldwide. In Mexico, the Ministry of Education (Secretaría de Educación Pública, [SEP]) has overseen the validation and certification of institutional programmes and courses since its foundation in 1921. Mexico mainly relies on national and international standardised tests for different purposes, such as secondary and tertiary admission in educational institutions (e.g., CENEVAL<sup>1</sup> or EGEL<sup>2</sup>). Similarly, the majority of private higher education institutions and some public higher education institutions demand their students to demonstrate the ability of comprehension and production of a second language to be awarded at the end of their degree (e.g., EXIT<sup>3</sup> or FCE<sup>4</sup>). These test-scores also could open doors to studying abroad or seeking job opportunities locally or internationally.

EFL teachers are responsible for delivering English language instruction in institutions (SEP, 2011); however, research studies have demonstrated that teachers have a lack of preparation in this field in Mexico (Davies, 2008; Gonzalez, 2018; SEP, 2018). A lot of attention has been given to private institutions since they prioritise foreign languages programme such as English and French, but the public sector seems to be affected by different external factors (Davis, 2008). In addition, research studies have also identified that there is a difference in relation to students' achievements among those who study in a public, private or language institute sectors (Moore, 2017). It is in the institutional context where teachers should be aware of how to deliver knowledge to students throughout different teaching techniques and practices that enable learners to build solid knowledge, which eventually might be tested. So, once the teachers have tested their students in relation to what they have learnt, teachers collect their evidence and students could be aware of the assessment procedures, communication of results and advantages of assessment towards improvement. Classroom assessments play a vital role in any educational system since it is inside the classroom where teachers are required to recognise different aspects

---

<sup>1</sup> **CENEVAL** (The National Centre for Evaluation of higher education) This institution works in collaboration with the Ministry of Education in Mexico to develop the National Measure to enter undergraduate and postgraduate courses. <http://www.ceneval.edu.mx>

<sup>2</sup> **EGEL** (*Examen General de Egreso de Licenciatura*) General Exam for Exit Purposes for Higher Education <http://www.ceneval.edu.mx/examenes-generales-de-egreso>

<sup>3</sup> **EXIT** (*Exámen de Inglés de Tamaulipas*) English Language Test in Tamaulipas (Universidad Autónoma de Tamaulipas-Exam for Graduation Purposes (B1)

<sup>4</sup> **FCE** (First Certificate of English) Cambridge Main Suite: Exam at B2 (Common European Framework for Language Reference)

of assessments like developing, administrating, marking and interpreting test-takers' scores (Fulcher, 2012).

Literature suggests that there is also a lack of training in assessment among language educators in Latin American countries including Mexico and Colombia (Davis, 2008; Giraldo, 2018a). Professional development opportunities have assisted language teachers by equipping them with a broader understanding of different methods for measuring students' language proficiency among educational contexts, and to highlight what teachers require to understand and to put in practice at the moment of conducting language tests and other assessment forms (Malone, 2009; Fulcher, 2012). However, many of these professional development courses do not meet the teachers' current needs to successfully achieve higher standards in education (OECD, 2010).

In term of language education, SEP has developed and approved several educational reforms to promote better opportunities for language learners, teachers and school authorities. Most of these efforts have primarily benefitted primary and secondary education institutions. The National English Programme in Basic Education (NEPBE) started in 2006 and stated that the English language should be part of the curricula from primary education (6 to 12 years old) to secondary school (13 to 15 years). In 2008, SEP redeveloped NEPBE to create the National Strategy to incorporate English language as a subject in primary and secondary schools for public and private sectors.

The Ministry of Education has inevitably the responsibility to collect students and teachers' learning evidences through school-based and external assessments, such as national measures for admission purposes towards the next learning stage such as upper-secondary or tertiary educational programmes. Some examples of high-stakes tests for English language proficiency are the CENNI<sup>5</sup> (Mexican Scale of English), IELTS<sup>6</sup> and TOEFL<sup>7</sup>. These latter two English tests are administered around the world to assess stakeholders' language ability across different language skills. These tests have multiple uses, such as assessing the language skills for pursuing a degree in an English-speaking university, seeking business opportunities at a worldwide scale, or for migration purposes. However, many test-takers do not meet the language level required and they are required to retake these examinations in the future. Due to the lack of success in obtaining the expected results from these examinations, educational institutions and governments have

---

<sup>5</sup> **CENNI** (National Certificate of Language Proficiency. This is the Mexican Scale of English which has been developing since 2013.

<sup>6</sup> **IELTS** (International English Language Testing System) is a test developed by Cambridge English Assessment, the British Council and the IDP Australia.

<sup>7</sup> **TOEFL** (Test of English as a Foreign Language) is a test developed by the Educational Testing Services

## Chapter 1

taken matters seriously in this regard by providing and supporting teachers to deliver high-quality teaching and learning practices, in order to boost learners' knowledge. As a result, teachers ought to be responsible for providing different strategies and techniques to their students to be able to meet those educational standards; however, research studies also demonstrate that this is still a gap to obtain these outcomes. Teachers play the fundamental role on how students perform; therefore, it is vital to examine how teachers self-perceive their assessment skills and practices in their classes.

This current study aimed at exploring EFL teachers' (from upper-secondary education and higher education institutions) educational sectors (private, public and language institutes) in Tamaulipas, Mexico and examining their assessment skills and practices, as well as to scrutinise how these these gain professional development. This research study has five main aims:

- 1) To *examine* the status quo of language assessment literacy in Tamaulipas,
- 2) to *explore* how EFL teachers self-perceived their language assessment skills and practices (TLASP),
- 3) to *scrutinise* EFL teachers educational sectors and discover discrepancies, if any, among their skills and practices
- 4) to *identify* any relationships between EFL professional development and language assessment and to *examine* which areas of language assessment require EFL teachers to empower their language assessment literacy
- 5) to *measure* the impact of a workshop for teachers' language assessment literacy.

With the intention to examine and explore the current situation and educational trends in EFL teachers and language programme managers with respect to their assessment skills and practices in the State of Tamaulipas. This research encompasses the '*status quo*' which refers to: a) test development and responsibilities, b) test relevance and occurrences, c) assessment of language skills and linguistic components, d) assessment experiences, e) professional development and f) critical assessment issues in teachers' institution. The following section introduces the research questions that were answered for the study.

### 1.4.1 Research Questions

In order to meet these abovementioned research aims, I have addressed five research questions for this study:

- 1) What is the *status quo* of EFL teachers' language assessment literacy in Tamaulipas?

- 2) What are the discrepancies, if any, in EFL teachers' language assessment skills and practices among different educational sectors in Tamaulipas?
- 3) What are EFL teachers' and programme managers' backgrounds in their professional development towards language assessment literacy?
- 4) In what ways has the workshop influenced EFL teachers' awareness of language assessment literacy?
- 5) What was the impact of a workshop in language assessment literacy development on EFL teachers in Tamaulipas and their educational sectors?

## 1.5 Scope of the Study

This research study was carried out in Tamaulipas, a Mexican state located in the North-East region. For the purpose of this research, I decided to consider two stakeholders; language teachers and language programme managers due to the lack of conducting classroom observations. Therefore, language programme managers provided an expanded perspective on how assessments were taken place in their educational institutions. These participants were EFL teachers at upper-secondary (students' age from 16 to 18 years old) and tertiary education institutions (from 18 years old). Two research studies were conducted to maximise the outcomes of this research: exploratory and experimental. The exploratory research was conducted to scrutinise participants' backgrounds and their educational sectors in relation to their assessment skills and practices, as well as their professional development. Firstly, an adapted instrument developed by Zhang and Burry-Stock's (1997) '*Assessment Practices Inventory*' (API) was used to explore how EFL teachers and programme managers self-perceived their assessment skill and practices. Some of the original items had to be discarded since they did not meet the purpose of the study (e.g., parental support, young learners' assessments, etc.).

This previous survey was adapted and incorporated as one section of the final version of the instrument, which I named as '*Teachers' Assessment Literacy Development Inventory*' (TALDI). This research instrument followed a hybrid theoretical framework to scrutinise classroom assessment practices with the assistance of the assessment literacy standards developed by the AFT (1990), Xu and Brown' (2015) theoretical framework and Giraldo's (2018) language assessment literacy statements.

Aside from TALDI, I decided to conduct semi-structured interviews with Tamaulipas EFL teachers (n=12) and language programme managers (n=3) to analyse in-depth their educational sectors, and how their assessment skills and practices differed from each other. A thematic analysis was conducted to identify differences alongside the quantitative data in relation to

## Chapter 1

participants. The findings from the TALDI and interviews were analysed and triangulated to discover any similarities and/or discrepancies among participants' assessment skills and practices among the sectors. These previous outcomes were essential to plan, develop and implement a workshop as a professional development training course to empower teachers' assessment literacy.

Moreover, an experimental research was also conducted to further seek data to examine which areas of language assessment literacy were essential for the EFL teachers in Tamaulipas context. This workshop consisted of 10 sessions in which participants (n=10) from the three different educational sectors took part. The workshop lasted 20 hours and aimed at three main aspects: 1) to explore the fields of testing and assessment to have a better understanding on how teachers conduct these practices inside the classroom, 2) to increase awareness about the essential qualities from the usages of tests and assessment practices, and 3) to explore collaboratively among participants different critical issues as to when and how assessments took place at their institution.

Once the intervention was concluded, participants' materials were analysed based on their perceptions and progress. A measure was developed with the intention to build a construct of course contents, with this being administered before and after the workshop. In addition, personal reflective tasks, practical tasks and a course evaluation sheet were analysed to measure the participants' impact. Moreover, a focus group interview (n=6) was conducted several months later through a Skype call to gather further information to seek whether participants were able to implement the knowledge and practices acquired in the two-week course. The research scope of the study focused on the knowledge, skills and practices language teachers are required to understand to administer them in the classroom. The aspect of the use of technology for language testing and assessments was discarded since some teachers did not have these facilities in their institution. Nevertheless, participants mentioned that this factor was essential for some of their educational sectors. The following section presents a brief summary of the significance of this research towards theoretical, practicum and professional contributions.

### **1.6 Significance of the Study**

The findings from this research study could bring several key benefits to understand the fundamental aspect of language assessment literacy that relies on teachers' educational contexts. The context (where teachers perform the act of teaching) has been discussed in different research studies and literature; however, it should focus as the main priority due to individual characteristics that might affect teachers' practicum trajectories in relation to their assessment

skills and practices. Also, based on the research findings, I had developed a conceptual framework on how stakeholders from different educational sectors should classify assessment with regards to their institutional operational system. In addition, some descriptions based on the language assessment literacy standards were developed in relation to teachers' language assessment literacy development. The outcomes of this investigation could also be a call for teachers to provide guidance to use a variety of assessment skills and practices among educational sectors. Also, policymakers in the state of Tamaulipas educational system should develop functional and contextualised courses where teachers could share their ideas, reflect and evaluate on their practices.

## 1.7 Structure of the Thesis

This introductory Chapter 1 presented the background of the study, as well as the rationale for this. It introduced the research context, the language assessment literacy standards and participants' educational sectors. It then highlighted the research aims and questions for this study. Then, Chapter 2 provides a comprehensive review of the literature from the historical and general terms of assessment literacy for general education towards teachers' language assessment literacy. It highlights the standards and components of what teachers and programme managers require to know to maximise their assessment skills and practices. It also presents an analysis of how language assessment literacy could be influenced in the Mexican educational context and individual features are discussed (i.e., teachers' experiences, academic, practicum and professional trajectories, as well as the institutional operational systems). Chapter 3 presents the research design and methods that were used to answer the five research questions. It describes the process of development, two pilot studies and administration of the research instruments. It provides participants' profiles and the sample of each of the four phases of data collection. This chapter also presents the process of development of the workshop and the different tasks that participants had to achieve. Lastly, this Chapter provides the ethical considerations for participants and concludes with a relevant summary for understanding the findings of the research questions.

Chapter 4 presents the findings of the survey (TALDI). It details the background of the participants in relation to their *status quo* of teachers' assessment skills and practices. It reports on how the participants differ from their professional development trajectories, educational sectors and experiences. In addition, it reports on the analysis of variances (ANOVA) findings from the three educational sectors. Chapter 5 reports on the qualitative findings from the semi-structured interviews among language teachers and programme managers to answer research questions 1, 2 and 3. Chapter 6 presents the findings of the participants' performances

## Chapter 1

throughout the workshop. It reports on the measures to observe where participants made progress, the practical and reflective tasks that participants completed were reported. The workshop evaluation sheet and the focus group interview findings are also discussed. This chapter concludes with a summary of the relevant findings to research question 4 and 5.

Finally, Chapter 7 presents the discussions of the findings and how these were compared to other research studies, with a comprehensive revision of the research questions. Closing remarks of this research study and how this research could be further conducted.

## Chapter 2 Language Assessment Literacy

This Chapter provides a review of the literature regarding vital aspects for the purpose of this research study. This begins with an overview of the terms ‘assessment’ and ‘literacy’. It presents a comprehensive approach to understand how the term literacy has evolved from its 20<sup>th</sup> century connotation to its practical application in different domains. Also, it includes three main purposes of assessments and some definitions which are included to highlight the foundations of the concepts of tests, assessment and evaluation. Section 2.2 introduces the term language assessment literacy and how this term has changed over the last three decades. Then, a section (2.3) has been included to highlight whom requires assessment literacy and those five dimensions for measuring stakeholders’ levels. Section 2.4 presents the teachers’ language assessment skills and practices (TLASP) with an overview of the standards chosen for the purpose of this study, as well as three matrices of TLASP for the knowledge of assessment, technical considerations and school-based language assessments. Section 2.5 presents a multilayer analysis of the Mexican assessment policies in education which are divided into six different levels from the evolution of assessment in Mexico to teachers’ individual perspectives and their professional development backgrounds. Section 2.6 reviews empirical research studies in language assessment literacy. This Chapter ends with a summary of the most relevant aspects of teachers’ language assessment literacy and referring to the literature gaps and their research questions.

### 2.1 Literacy and Assessment

According to the Oxford English Dictionary, literacy mainly refers to the competencies of reading and writing communication, as well as numerical ability. This term has been frequently applied in basic educational programmes to equip pupils with essential tools for their communication and basic numerical understanding. Several authors have recognised that the word ‘literacy’ has commonly gained and raised much importance across professional, scientific and academic domains. Researchers such as Taylor (2013) and Dunkley (2014) have pointed out that literacy or literacies are those abilities, skills and techniques that make someone proficient in a discipline. In the same way, Brumfit (2010) has also suggested that this word has changed its connotation due to the social impact and professional requirements for individuals to become skilled in this globalised and competitive world. As an example, ‘*technological literacy*’ refers to the core of knowledge, skills and practices of the use of technological devices. In the field of language education, the Alberta Education Institute (AEI) defines literacy as “*the ability, confidence and willingness to engage with language to acquire, construct and communicate meaning in all aspects of daily living*” (AEI, 2019). This institution regards literacy as an individual’s set of skills to use language in different situations in a particular context. Therefore, we can

## Chapter 2

understand that literacy is used to refer that someone is equipped with the required knowledge, skills and techniques that makes someone competent in a particular discipline. A significant number of literacies can guide individuals to understand one or several domains by underpinning vital knowledge, abilities and practices.

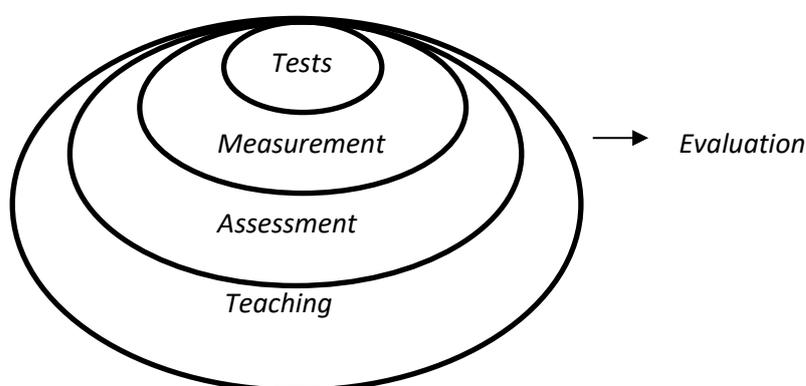
Since this literature review is based on assessment in education, it is vital to review fundamental elements to become assessment literates. It is relevant to identify the purposes and reasons why assessments are used. According to Fulcher (2012) assessments are not only utilised for educational purposes, but these have the potential to go beyond the classrooms' walls. Fulcher listed three main reasons for using tests and assessments, the first is due to the purpose of accountability. Those assessments that have been stipulated by policymakers, governments, and national and local authorities. These could be internal or external examinations that deal with the purpose of hiring new personnel for a work position or national examinations for entering a new educational level. Brindley (2008) considers these examinations as political tools since the results of these have a specific impact within a society. As a clear example, the Programme for International Student Assessment (PISA) evaluates students' competence in the first language, mathematics and social sciences for those students who are concluding primary and secondary education programmes across the globe. These examinations have a global impact as governments use this testing tool to seek better national positions for educational levels and receive recommendations on how to improve the local, regional and national educational standards (PISA, 2012).

The second reason deals with mobility. Different stakeholders required assessments to demonstrate their language ability to move overseas for seeking better job and/or educational opportunities. These assessments could also be a prerequisite to promote those national and international students' exchange programmes. Hogan-Brun, et al. (2009) argued that language educators are not directly linked to this purpose of assessments, but they could be considered fundamental pillars for language preparation courses to empower test-takers' learning pathways. Also, language educators, especially those who teach a foreign language that is not their mother tongue, are required to demonstrate a certain level of proficiency and be qualified to be able to teach the language to perform as language teachers (Kunnan, 2009; McNamara, 2008; Extra et al., 2009; Slade & Möllering, 2010). In the Mexican context, the use of these tests provide better opportunities for travelling to other countries and to find better opportunities for jobs in or outside the country (INEE, 2005).

The third reason relates to educational systems and takes place in a classroom setting. Fulcher (2012) states that teachers could be able to understand the methods for transferring

knowledge into assessments; in other words, the ability to align the construct of a test to the learning objectives of a course. Teachers are mainly responsible for designing, developing, administering and marking of language tests, as well as other assessments as part of their teaching requirements (Stiggins, 1991; Fulcher, 2012; Scarino, 2013). Learners are required to complete a diverse selection of tasks and activities defined by their teachers, to learn and boost their knowledge to empower their learning trajectories (Black & William, 1998; Rea-Dickins, 2006, 2008). Although the impact of these assessments might not be as significant as the prior reasons for assessments, accountability and mobility, classroom tests could bring satisfactory opportunities for learners to shape their thinking and ways to seek improvement in any discipline and, as well as to expand their experience in completing these tools satisfactory (Cheng, 2008; Wall 2005, 2012). Therefore, tests allow students to identify the areas where they need to improve to meet those requirements and learning objectives of their course.

Hence, to become an assessment literate, stakeholders (e.g., learners, teachers, programme managers, directors, etc.) require understanding why assessments are relevant and highlight the implications they could provide to different aspects in education or professional life. It is also fundamental to understand some definitions related to the use of assessments. Literature has suggested that sometimes teachers misunderstand certain features of assessment in their teaching practice (e.g., tests, measurement, assessments, and evaluation). Although these four words have interchangeable properties, each of them owns unique qualities. Several researchers have suggested that teachers should be able to differentiate these terminologies (Bachman, 1990; Bachman and Palmer, 1996; McNamera, 2000; Brown and Abeywickrama, 2010; O’Sullivan, 2017). Brown and Abeywickrama defined these definitions with an illustration in their book. Figure 1 displays how these authors differed from testing, assessment, measurement, and evaluation in relation to the practice in the classroom.



**Figure 1 Brown and Abeywickrama’s (2010:6) Diagram**

Figure 1 includes four circles: starting at the outer-most circle, *Teaching* deals with the teachers and sometimes with other stakeholders who deal with the instruction of contents and

## Chapter 2

learning objectives in a particular educational course. These contents that are covered in the classroom should reflect the institutional decisions and their teaching context. As an example, for language education, language teachers need to be aware of developing proper learning outcomes that would be based on the level of proficiency from the students. The next inner circle, *Assessment* deals with the process of developing, marking and grading testing instruments and other uses of alternative assessments, as well as providing students with relevant feedback and categorising the test takers based on their performances. It also focuses on the process of collecting, marking and analysing learning and teaching evidences, not necessarily related to a test, but also to the test-takers' personal characteristics, levels of progression and students' achievements (Thornbury, 2006; Jones, Saville & Salamoura, 2016). Brown and Abeywickrama (2010) with agreement with O'Sullivan (2017) defined assessment as a developmental process to boost learning; in other words, the way teachers use different assessment methods to maximise students learning. These methods could be the use of worksheets, quizzes or rubrics, so as learners can develop their knowledge, skills and practices using a specific domain or discipline.

Another important aspect to assessment literacy, in my view the most relevant at the moment of comparing national and international scales assessments, is *measurement* or measure which provides a quantity (grade or mark) to identify how test takers have performed based on a scale or criterion. Bachman (1996) states that measurement is the process of allocating numbers, not only to what the test takers know, but also to several attributes towards abilities or knowledge demonstrated. This measurement is defined by the local, regional or national educational authorities with the intention to determine how students' progress. In Mexico as in Tamaulipas, the measures for education, not only language education are based on a scale from 0 to 10, based on students' performance. The Mexican Ministry of Education has developed a scale with a minimum pass of 6 referring to satisfactory and 10 referring to an excellent performance. This scale is widely use in primary, secondary and higher educational institutions (Ministry of Education SEP, 2012).

The innermost circle, *Testing or Test*, refers to the designing and development of instruments that teachers tend to administer to their learners with the intention of identifying and quantifying the amount of knowledge they have learnt. The purpose of these tools varies based on what teachers, school authorities or policymakers have required. Instruments as tests, checklists or to-do lists can measure knowledge, skills, aptitudes, attitudes and learning performance progression patterns, not solely for learners, but also for teachers. By the same token, O'Sullivan (2017) defines tests as tools for making judgmental decisions in which teachers and other stakeholders, such as language programme managers, make educational decisions such as failing or upgrading students to the next learning stage.

Outside of the circles, we have *Evaluation* which focuses on the critical and judgemental analysis of the instruments' outcomes to be able to scrutinise certain aspects that could help teachers, institutional authorities and policy makers to revise, amend and improve their teaching, learning and assessing methods. It requires the test scores and test-takers' performances to judge the progression analytically, and this provides implications to improve the syllabi, curricula and educational systems (Thorndike & Hagen, 1961). Identifying and comprehending these five important concepts of educational assessment, teachers could develop better opportunities to grow institutional assessments practices towards the improvement of educational practices. It also guides stakeholders to understand the needs for testing and assessment, not only for educational purposes, but as vital components for social perspectives (McNamara & Roever, 2006).

The focus of this sub-section was to explore and mainly review the mainly third reason for assessments that Fulcher suggested: classroom practice. Tests and assessments are considered two vital areas for educational and social purposes; however, it is in the classroom where teachers and learners are frequently involved in learning, teaching and assessing language. As a result, to develop assessment literacy, these fundamental definitions allow teachers to highlight the purposes and uses of tests and assessments to maximise their time for development and administration, as well as communicating results and finding opportunities to expand knowledge. The following section introduces the term of language assessment literacy and provides a chronological review of this term from the last three decades.

## 2.2 Language Assessment Literacy

Historically speaking, the term Assessment Literacy was firstly coined by Richard Stiggins almost three decades ago in 1991. He highlighted that teachers were not well-prepared with educational assessment knowledge; therefore, a call to improve teachers' assessment knowledge, skills and practices was made from different educational institutions, governments and organisations from several parts of the globe. This researcher stated that educators with assessment literacy should be able to differ and recognise '*what to assess, why to assess, how to assess, in order to examine the possible educational assessment critical issues and how to prevent them from occurring*' (Stiggins, 1991:4). This was not, however, the first researcher who pointed out that there was a lack of knowledge to conduct educational assessments adequately.

In the field of language education, Robert Lado (1961) indicated that teachers should have enough assessment knowledge and practice to accurately assess their language learners. Almost three decades later, Nunan (1988) identified a lack of assessment knowledge and the inefficient use of testing and methodologies in second language classrooms. Not long after Stiggins'

## Chapter 2

definition of assessment literacy in the early 1990s, several researchers have investigated different aspects of assessment literacy in different countries, as well as in various educational contexts. Research studies in assessment literacy have had implications in teachers and researchers to work in constant collaboration to identify what teachers require to maximise and bring those benefits to the educational process of learning and teaching.

Almost three decades had passed since assessment literacy was established as a concept for educators, researchers and testers, when DeLuca, et al. (2010) suggested that there has been three main movements of assessment literacy, fundamentally one per decade. The focus of assessment literacy in the 1990's was to identify what teachers should know and to prepare them to understand the importance of accurate knowledge of educational assessment in a broader orientation. This first period of time was since the development of the first standards of assessment literacy in 1990 from The American Federation of Teachers (AFT), National Council on Measurement in Education (NCME) and the National Education Association (NEA) developed *the standards for teaching competence in educational assessment of students*. These standards were 1) selecting assessments, 2) developing assessments for the classroom, 3) administering and scoring tests, 4) using scores to aid instructional decisions, 5) communicating results to stakeholders, 6) being aware of inappropriate and 7) ethical uses of tests (AFT, NCME, NEA, 1990). At that point in time, researchers and educational departments aimed at standardising what teachers should know about testing and assessment towards large scale or standardising measures, including summative assessments and teachers' development towards the usefulness of tests by analysing them psychometrically.

Stiggins' (1991) study reporting on that the abovementioned competencies were not enough for preparing educators for the use and knowledge of language testing and assessment. He stated '*We are a nation of assessment illiterates*' (Stiggins, 1991: 535). In fact, he added more abilities based on what language educators needed to understand the process of assessment clearly, as well as to make this easier to other stakeholders such as language learners. Towards the end of 1990's, the focus of assessment literacy had a significant impact on teachers to boost their assessment literacy levels to achieve the national standards. Brookhart (1998) highlighted that assessment literacy could assist other functions than only the assessments in the institution. Along the way, Stiggins (1999) indicated that the quality of the assessments should mirror the quality of instruction.

The second period covers the years between 2000 and 2009. One of the main changes of assessment literacy was oriented towards assessment for learning and assessment as educational support. In other words, those assessments that help students with their learning trajectories, but

also to assist teachers in interpreting scores to modify teaching and learning practices. This decade brought into the table several definitions of assessment literacy, as well as the initiation of the concept language assessment literacy. Paterno (2001) suggested that if educators possess a deep repertoire of assessment and testing competencies including terminology, use of statistical and psychometrical analysis, and familiarity to test development could empower their educational settings. Mertler and Campbell (2005) pointed out that educators required several abilities to develop tools to assess learners and facilitate learning. Similarly, Stiggins (2002) and Popham (2004) suggested that assessment literacy relies on the principles of testing and assessment such as validity and reliability. It was also in this decade when assessment and pedagogy joined forces, as a result the process of learning and assessment could be enhanced.

Then, Boyles (2005) and Hoyte (2005) define assessment literacy as a “toolbox” that includes the same elements as Stiggins’ initiate movement with a combination of theoretical and practical elements for constructing and administering testing instruments. In other words, language assessment literacy encompasses the critical and practical ability of the appropriate use of assessment and language teaching based on the contexts where these occur (Johnson & Kress, 2003 and Leung and Read-Dickins, 2007). Likewise, Mertler and Campbell (2005) and Otero (2006) claimed that teachers need to be trained and equipped with strategic elements such as skills, pedagogical and systematic knowledge of assessment.

Another significant definition of assessment literacy was the one provided by Volante and Fazio (2007) that assessment literacy enables stakeholders to become aware of their situations where assessment takes place and how educators can manage the use of instruments for their development. Davies (2008) pointed out that language assessment literacy relies upon three main components; *skills* (e.g., item-writing, development of materials, etc.), *principles* (concepts about testing principles, ethics, etc.), and *knowledge* (e.g., theories of test development, measurement, language education, etc.). By the same token, Inbar-Lourie defines this concept as *‘having the capacity to ask and answer critical questions about the purpose of 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’* (2008: 389). Language teachers need to be able to understand how the properties of language assessment work with regards to improving their learning process. Inbar-Lourie (2009) highlighted the importance of language assessment literacy for the role of decision-making in various areas where assessment takes place or becomes a priority for the trajectories of teachers and students’ performances. Davison and Leung (2009) agreed that teachers with a higher level of understanding and practicality of the use of assessments can increase the quality of the internal and external functions of the teachers’ performances.

## Chapter 2

The third period of assessment literacy consisted from 2010 to the present time in which one of the aims of assessment literacy was that educators could gather information about test performances, understand the purpose of testing, and to complement with instructional pieces of evidence so as to improve their educational systems (Siegal and Wissehr 2011; O'Loughlin, 2013). The focus on assessment for learning and assessment as educational support are still gaining power. Scarino (2013) not only defined assessment literacy as a set of skills or competencies, but also as an enabler of awareness to explore and understand how assessment is considered and treated as part of the educational purposes. Willis, et al., (2013) pointed out that assessment literacy is a dynamic and complex process that includes their culture and teachers' needs of improvement, not only for educational purposes, but that seeks to place a considerable importance to increase the status of the context. They pointed out that language assessment literacy dimensions rely on how the teacher is involved with a vast number of classroom techniques, skills and practices, as well as knowledge of theory of language acquisition and assessment. Lam (2015) defines assessment literacy as the acquisition of knowledge, skills, and principles of testing as well as the use of assessment in the language classrooms. By the same token, a recent definition of Assessment Literacy by Fulcher cited in Tsagari (2014):

*'The knowledge, skills and abilities 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: 125).

Fulcher's (2012) definition of language assessment literacy is, to some extent, the most fulfilled that covers three levels of assessment impact: individuals, institutional and society perspectives in assessment. This researcher made full emphasis that assessment is part of a cultural perspective and it needs to be explored per context. To this point, researchers have come up with the definitions of language assessment literacy. The importance of a high level of knowledge in the variety of types, uses and purposes of assessments in different contexts could provide stakeholders diverse views and perspectives about the appropriateness of this field (Pill & Harding, 2013). Assessment literacy enables teachers to comprehend the significance of the testing and assessment processes.

Language educators need to understand why assessment is a key factor that can contribute to a better performance of the teachers, learners, but more importantly, it can increase the

profile of an institution (Eames, 2003). Likewise, some researchers have different perspectives on what are the essential components to increase the profile of an assessment literate. Llosa (2011) indicated that to improve the quality of education, *'it is crucial to develop rigorous standards and align instruction and assessment'* (Llosa, 2011:367). As a result, Taylor (2013) developed a domain of competencies that assessment literates need to understand and apply into their teaching and assessing practice. Therefore, she pointed out the following elements; 1) technical skills, 2) scores and decision making, 3) language pedagogy, 4) local practices, 5) knowledge of educational assessment theories, 6) assessment and testing principals and concepts, 7) socio-cultural values, and 8) personal beliefs and attitudes. Harding and Kremmel (2016) revised these elements to expand the level of knowledge and understanding of testing, assessment and evaluation. They proposed revised domain on the following aspects: 1) theoretical knowledge about language and language learning, 2) technical skills: language assessment construction, administration, scoring and evaluation, and 3) the impact in society of tests and scores.

A research contribution from a DeLuca, LaPointe-McEwan and Luhanga's (2010) study was to analyse a large number of language assessment literacy standards and principles across different English speaking countries (e.g., United Kingdom, The United States of America, Australia, New Zealand). They conducted a document analysis and a psychometric analysis of several instruments. They reviewed 36 instruments that have been used to measure assessment literacy. Based on their thematic analysis, they listed eight main standards that teachers should take into consideration: 1) assessment purposes, 2) assessment processes, 3) communication of assessment results, 4) assessment fairness, 5) assessment ethics, 6) measurement theory, 7) assessment for learning and 8) education and support for teachers (DeLuca, et al. 2010: 255). Two standards were modified from the AFT 1990, the assessment for learning and assessment education and support for teachers. Fulcher (2012), Vogt and Tsagari (2014) and Taylor (2009) criticised DeLuca's study due to the lack of the social component in language assessment literacy.

Xu and Brown (2016) reviewed a total of 100 empirical and non-empirical studies to create a framework of what language teachers require to develop assessment literacy. These authors highlighted that knowledge of assessment (assessment purposes, content and methods, grading, feedback, self- and peer- assessments, interpretation, communication and ethics), teachers conceptions of assessment, practicum (the perform of teaching and assessing language in the classroom) and professional developmet could enhance teachers' language assessment literacy with institutional methods of reflection, participation and collaboration with members of the educational system. The authors' framework had a multi-directional flow that makes, to some extent, it difficult in understanding the effects of each of the components. However, it brings

many benefits to understand how the context places a crucial aspect in the way that teachers and other stakeholders perceive assessment.

Under the context of Latin-American education, Giraldo (2018) developed a framework with three main components: principles, knowledge and skills. Principles links to the awareness on how assessment is treated; knowledge are theoretical applications of applied linguistics, educational assessment and the context where assessment occurs; and skills that lead to the learning, teaching and assessing language practices. Although this study was written based on researchers, teachers and researcher's personal experiences, this could bring a proper understanding in categorising teachers' assessment skills and practices. As this researcher noted in the checklist he developed, teachers in pre-service and in-service could use these descriptors to highlight their needs.

Researchers have identified language assessment literacy as a need for language teaching professional development, standardisation of materials for pre-service and in-service teachers, development of the areas of test construction and the socio-cultural perspectives as well as socio-political elements to improve the language testing in the society and educational context. Inbar-Lourie (2013) points out that the need of a threshold level is essential for stakeholders who are close related to assessment and language. This threshold contains four levels divided into how knowledgeable the stakeholders require to have: 1) general assessment, 2) language issues, 3) language assessment, and 4) the context. Assessment also deals with the social perspective and the importance of educational based on the context. Inbar-Lourie (2009) makes a distinguish line between the testing and assessment cultures. The testing culture is the psychometric perspectives on how tests are constructed, validated and used. Some nations believe that that assessment holds an important aspect of educational purposes throughout the administration of tests. On the other hand, there is the assessment culture, which requires understanding on how learners and teachers interact through different assessment tools and methods. The assessment culture requires understanding on how learners acquire the vital knowledge through their social and cognitive processes. The following section describes the stakeholders those who require assessment literacy and how to distinguish them based on five different dimensions.

### **2.3 Language Assessment Literacy: Stakeholders and Dimensions**

Several studies have tried to answer the question *who needs assessment literacy*. As has been said, language assessment takes an important role in the language classroom since teachers and students are constantly involved in this educational practice (O'Sullivan & Chalnack, 1991). Nevertheless, these are not exclusively working with this educational practice, in fact, in some institutions, language programme managers, test developers, parents and other stakeholders are

also being influenced by test scores and assessments (Taylor, 2013; O’Sullivan, 2017).

Researchers across the world have pointed out that any individual who has taken a test needs assessment literacy to be able to understand the purpose and its impact.

Over the last two decades the number of people who are involved in assessment has increased due to the current educational and social aspects. Green (2013) listed several stakeholders: 1) policy makers, 2) government agencies and bureaucrats, 3) educationalists and employers, 4) media and general public, 5) test takers, and 6) teachers and teacher educators. Likewise, Berry and O’Sullivan (2016) categorised stakeholders based on three sections: 1) developers, 2) direct users, and 3) indirect users. Table 1 shows the stakeholders based on Berry and O’Sullivan’s classification.

**Table 1 Berry & O’Sullivan’s (2016) Stakeholders Classification**

<b>Developers</b>	<b>Direct Users</b>	<b>Indirect Users</b>
Designers	Test takers(students)	Policy Makers
Item writers	Teachers	Employers
Task makers	Parents	Educational administrators
Analysts	Admission officers	Training providers
Academics	Delivery staff	

The development of instruments to track stakeholders’ assessment literacy to measure their understanding and application of assessments in their contexts, have made a significant important contribution to classify the stakeholders’ language assessment literacy levels. Kaiser and Willander (2005) introduced five stages of assessment literacy based on assessment knowledge and practices. Table 2 describes the level of literacy and defines each stage.

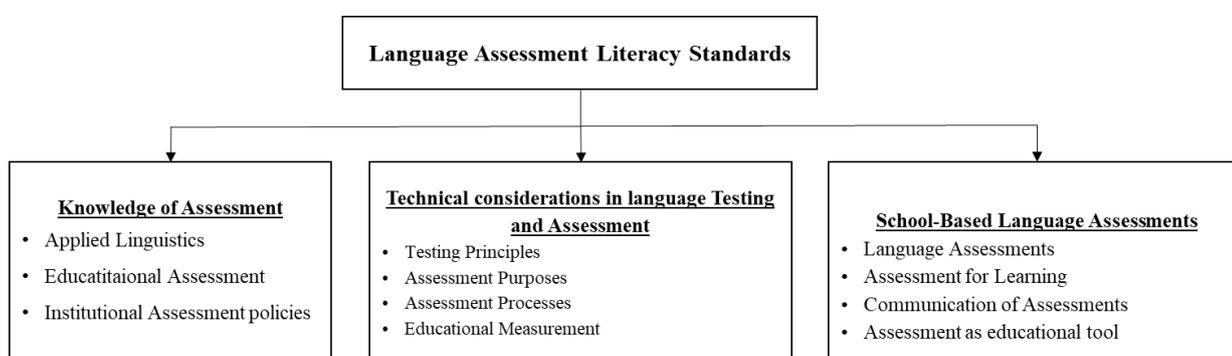
**Table 2 Levels of Assessment Literacy**

<b>Stage of Literacy</b>	<b>Definition</b>
Illiteracy	Unfamiliarity of language assessment concepts and methods.
Nominal literacy	Understanding of specific concepts but indicates misconceptions.
Functional literacy	Sound understanding of basic assessment concepts.
Procedure and conceptual literacy	Understanding central concepts of the field and using knowledge into practice.
Multidimensional literacy	Knowledge extending beyond ordinary concepts including: philosophical historical and social dimensions in assessment.

For the purpose of this literature review, language teachers and programme managers would only be taken into consideration. At the classroom level, students and teachers are mostly direct in the practice of assessment. Nevertheless, at the institutional perspective, language programme managers also take a large responsibility for the process of development, pilot and administration of tests, since several teachers do not prepare nor plan their language tests.

## 2.4 Teachers' Language Assessment Skills and Practices (TLASP)

This section introduces the teachers' language assessment skills and practices (TLASP) which focuses on three main components or areas of language assessment. Taking into consideration the three abovementioned sets of different language assessment literacy standards (AFT, NCME, NEA, 1990; Xu & Brown, 2016; Giraldo, 2018), it was vital to categorise them in these three main components of this field: knowledge of assessment, technical considerations in language testing and assessment and school-based language assessments. Figure 2 illustrates the standards of language assessment literacy under the umbrella of the three components following Davies (2008) classification of language assessment literacy. These three components and their standards are described in the following sections and introduced with a comprehensive review of literature highlighting these aspects to enhance teachers' language assessment literacy in the language classroom.



**Figure 2 Language Assessment Literacy Standards for the Language Classroom**

### 2.4.1 Knowledge of Assessment

This section of the conceptual framework comprises three standards: knowledge about applied linguistics, educational assessment, and institutional assessments that deals with the national and local educational contexts. In order to explore and examine the field of language testing and assessment, more precisely, the aspects in which the language is being used, it is important to highlight how language is being developed, stored and applied within a context. It is worth keeping in mind that to develop teachers' assessment skill and practices, researchers and practitioners are required to consider understanding the field of applied linguistics and epistemological assumptions to scrutinise the foundations to empower language learning in education throughout assessments. In addition, examining the social context where teachers can acquire and develop knowledge to be able to use a diverse repertory of assessment techniques and skills. Moore (1993) indicated three possible types of knowledge that learners and teachers require to highlight to be aware of their progress and their learning trajectories. Declarative, procedural and conditional knowledge are required to maximise the progress patterns in language

learning. Declarative knowledge (knowing what) focuses on how learners have made any significant progress in certain skills or techniques to achieve a task. This can be tracked by the test-takers' scores in a period of time or the number of hours being instructed. Procedural knowledge (knowing how) deals with the process of what skills and techniques need to be used in order to achieve a task. For language education, this can be perceived by how teachers and learners interact with a diverse of strategies to comprehend language better. Finally, conditional knowledge (knowing when) focuses on how test takers need to use certain skills or techniques that a task requires a setting or context, where language can be used and performed. These three types of knowledge need to be input by the teachers through different tasks, teaching approaches and learning approaches towards educational assessment. Constant testing and assessments could be considered a learning evidence, these may demonstrate that a teaching or learning aim has been met. Nevertheless, it is highly important for teachers, as well as for students, to understand appropriate knowledge when the language is used and applied in different settings. Applied linguistics is a discipline that focuses on how the language is applied into a society of a context (Wallace 1991; Brown 2007).

#### **2.4.1.1 Applied Linguistics**

Xu & Brown (2016) indicated that knowledge in applied linguistics could benefit the levels of language assessment literacy. Similarly, Giraldo (2018) also pointed out that knowledge of the main features from linguistics were vital to promote better assessment practices in the classroom. Therefore, teachers are expected to know what language to use, how the language is formed and when language should be used depending on the context in which it is applied. To begin with, language teachers should understand the foundations of linguistics because teachers are frequently teaching these structures from the language which they are expected to differentiate from five main levels (Brown, 2007). These levels are briefly described as *Morphology*: which is the study of the words and how these can be modified by adding affixes to use and classify words into parts of speech (e.g., nouns, verbs, adjectives, etc....). As language teachers, we need to distinguish words from their parts of speech: verbs, nouns, adjectives and adverbs are few categories that have a vital function in the language. The next level is *Syntax*, which focuses on how to group words coherently to produce meaningful sentences. Then, *Phonology*, which focuses on how the words are pronounced along with phonemes, the minimal particle of sound, therefore, a sentence. *Semantics* is the study of the how the meaning of objects towards the use of the language and study of connotations. Finally, *Pragmatics* is the discipline in charge of the meaning, sense and application of words and sentence in an accurate perspective (Yule, 1996; Widdowson, 1996; Scovel, 1998). These five levels of applied linguistics are vital since these are the ones teachers are expected to be assessed. Therefore, comprehensive knowledge of applied

linguistics would allow teachers to be aware of students' performances and allow students to be guided through teachers' feedback (Xu & Brown, 2016).

The difference between the terms *learning* and *acquiring* are quite simple now. Learning is the action where cognition occurs. Most of the time, students are in the classroom because they need to learn new things. The students receive knowledge in form of instruction, which students process the information and negotiate the meaning to be able to produce an utterance. On the other hand, acquiring a language deals with the emerging process on how we listen to a language from a specific group of people or from a region. This refers to the study in which people acquire an additional language (Scovel, 1996). This is one of the most researched fields under the umbrella of linguistics. An invaluable field with the most significant voices from philosophers, linguists and educationalists. Some of the most relevant theories are the ones Krashen's hypothesised: The input hypothesis, the natural order hypothesis, the monitor, acquisition-learning hypothesis and the affective filter. All these hypotheses have been well studied. The term learning a language and acquire the language are, for some, easily to be misunderstood (Krashen, 1987, 1988). At this point, children of an early age are not considered learners, rather they are able to mimic or imitate what their parents do, say and how they act in certain situations. These images, the movement of the mouth, the sound of the voice, required to be processed by imitation and then this social interaction with their parents, children could be able to start making similar noises. This differentiation is important for the field of language assessment literacy. The rationale for this is because the assessment of language relies on the way that language was acquired or learnt (Lyons 1981; Thornbury, 2006; Popham, 2011; Tsagari, 2014). The following section provides the most relevant aspects of educational assessment like educational theories, assessment concepts and the assessments of students' progressions.

### **2.4.1.2 Educational Assessment**

After understanding how the language is being used and applied in different situations, it is important to identify what teachers expect at the moment of instruction in a classroom. This section includes educational theories to lay the foundations of the use of assessments to evaluate students' knowledge, assessment theories also play an important role for understanding educational assessment. This standard, educational assessment, deals with educational movements such as the Behaviourism, Constructivism and Socio-cultural Theory. In addition, features of assessment for/as/of learning complement the students' profiles at the moment of learning a language.

#### 2.4.1.2.1 Educational Theories

Cognitive research studies have gained a lot of attention since the decade of 1940 when psychologists and educational researchers such as Jean Piaget, Lev Vygotsky and Bruner suggested that cognition occurred when the individual reflects on their prior experiences and knowledge, with the intention to identify what already is known and to reach the unknown (Ellis & Robinson, 2008). Several theories were developed based on these researchers' philosophies. The ambition to understand epistemological and ontological perspectives of language testing and assessment, it is essential to comprehend the learning theories behind language education in relation to educational assessment. This section deliberates aspects to understand how cognition or, in other words, the process of learning is understood, acquired and applied. It presents educational researchers' perspectives and discusses the process of learning, from a constructivist to a sociocultural approach.

To begin with, constructivism refers to the learning process that is being developed by ourselves, in other words, constructivism deals with the epistemological, pedagogical and psychological features of the learning process is being constructed by previous experiences and by the identification of previous and new knowledge. This educational movement was more related to the social integration of agents such as parents, teachers and other students (Ellis, 1995). Constructivism theory was the result of the behaviourist theory, in which the learning process was controlled by the teacher. The behaviourism trend emphasised on how the learners receive knowledge by different techniques such as memorisation and drills. This previous learning theory contradicts the constructivist theory in which cognition occurs when the learner becomes active in the classroom and receives enough guidance and support to develop and understand our own knowledge of reality. Piaget and Bruner indicated that learners are individuals and each of us have a unique repertoire of experiences that are mentally organised to represent a reality that is being interpreted with what has been learnt and what it is further required to learn (Vygotsky, 1978, Kanselaar, 2012). We can say that learning is personalised by each individual and it is within individual where they inquire about new knowledge. Borich and Tombari (2004) highlighted that learners are provided with essential tools to be able to guide them during the learning process. This guide is sometimes referred as the teacher who assist learners to achieve learners' learning objectives.

In terms of educational assessment, language tests had also been impacted on these theories, since tests and other methods of assessment were guided by these trends of education. Test constructs were simpler since repetition was, to some extent, a unique way of learning based on the behaviourism perspective. This clearly changed when the aspect of constructivism, which to some extent refers to assess the language based on one peculiar aspect of the integration of

agents to develop, guided and assess what a learner could perform (Davis and Pearse, 2000). Nevertheless, it is essential to understand how the language is used based on the socio-cultural aspect and how this trend influenced language testing and assessment.

#### **2.4.1.2.2 Social-Cultural Constructivism Theory (SCT)**

Since using the language includes a variety of aspects like the learning and acquisition, the language is clearly expected to be used in a society or with a group of people where the previous knowledge of what language to use when using it in a specific situation. As a result, a significant contribution to the theory of socio-cultural constructivism was the Vygotsky's pioneer work referring to the language acquisition process when occurs based on the social and cultural interaction (Nunan, 1992). This Vygotsky claimed that children could acquire knowledge with the assistance of their teachers and parents. His work was known as the socio-cultural theory. This theory relies on the development of knowledge through social interaction that enables learners' opportunities to acquire new knowledge (words and phrases) on a particular situation with a higher-experienced person among them. Vygotsky introduced the term mediated or mediator which refers to someone who has developed superior knowledge and assist others whose knowledge is being developed. In addition to this, Vygotsky also implemented four tools for the social constructivism: a) cognitive apprenticeship which deals with the integration of teachers and students to negotiate meaning, b) the scaffolding which involves the teacher's support and guidance to empower learning, c) tutoring deals with the social interaction from other students and superiors in terms of their knowledge; and finally, d) cooperative learning which involves the experience and development of working with other students (Frawley, 1989; Lowenthal & Muth, 2008).

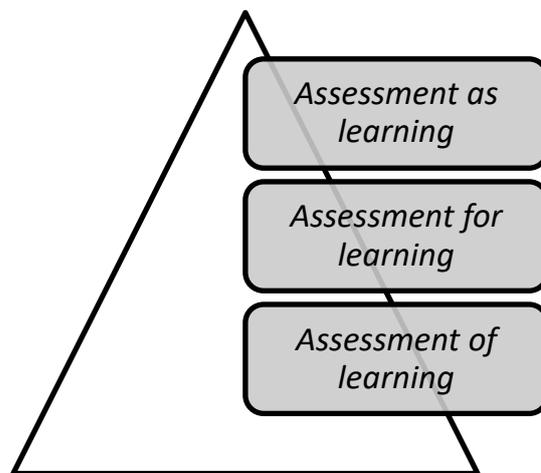
As an illustration, teacher education programmes are usually mediated by someone who has considerable experience in the field of education, these people are usually called trainers. This mediator assists teachers on their academic and professional trajectories by integrating a group of teachers into a course in professional development (Jones, et al., 2016). This process is known as assisted performance, where teacher trainers support their trainees by providing knowledge, practices and a repertoire of skills to become competent in their classrooms and therefore improve their teaching practices. In addition, the individual, in this case the teacher, allows communication with others from their kind. The purpose of this pre-service phase is to prepare new teachers to appropriately use the knowledge and techniques they had previously learnt to practice in their classrooms. This is a clear example how socio-cultural constructivism takes place. This stage is the appropriation of the foundations of knowledge which generally requires further input throughout teachers' practices and experiences throughout time.

The sociocultural constructive trends tend to rely on a collaborative understanding on how reality is seen and perceived by individuals taking into consideration social and cultural integration (Holliday, 1994). From this point of view, assessments have educational, social, cultural and psychological implications. As a result, assessments would mediate how students acquire knowledge. The facilitator or teacher, in this case, uses the test as the tools to collect and observe students' performances. An essential aspect of understanding assessments is to recognise how tests and assessments are seen from traditional and constructivist classrooms. For traditional classrooms, where learning materials are predominately course books and workbooks and the process of learning relies on repetition and translation, assessment is throughout testing resources and the scores are towards the summative evaluation. For constructivist classroom, where learning resources have been selected and adapted to facilitate learning and where the cognitive process relies of the support and dialogue with peers or teachers, assessment is viewed as the process of collecting learning evidences so as to the leaners become autonomous and aware of his learning process.

Educational assessment has taken a new role of importance for educational purposes. Theoretical and practical knowledge in assessment is essential to improve the quality of these and testing instruments, but at the same time, they tend to assist more towards the improvement of the quality of teaching (Tzagari, 2004, 2017). Educational assessment is mostly divided into two main categories that teachers need to understand to properly promote learning. Formative and summative assessments are two concepts very commonly reviewed in the literature of language assessment and educational assessment. *Formative assessments* are those instruments that students take to identify their learning progress, and these ones are not reflected in their final marks; on the other hand, *summative assessments* are the ones that are vital identifying whether the learners upgrade to the next learning stage. Formative and summative assessments are crucial for the development of learning and teaching practices. Therefore, formative assessment takes an important role for teachers and learners. Teachers need to have a repertoire of competencies on how to treat formative assessment; knowledge from both features and how these differ from these assessment tools. Moreover, the quality and the attributions from how assessment is evaluated. Therefore, it is essential teachers understand this feature of language testing and classroom-based language assessment.

Recent contributions towards the field of assessment have made possible to incorporate one fruitful aspect from the behaviourism trend, the learner. The evaluation of the learner is fundamental, but not only from the one that its teacher provides, but also for how that learner reflects on and self-evaluates its own language learning journey. Researchers have mainly use three-level pyramid of assessment. Figure 3 illustrates a three layer pyramid. At the top of this

pyramid is *Assessment as learning* which occurs when students are responsible for their learning process through assessment practices. These types of assessment are to encourage learners to become responsible for their own learning through their assessments. It deals with students' reflection to identify their learning trajectories with the assistance of their teachers' feedback. This method has a range of skills that students need to apply for discovering their own preferred methods and techniques for better learning and it can also bring a beneficial part to teachers to try to amend their teaching practice. O'Sullivan (2017) has suggested that students' voices are important for the development of their learning, but also to let teachers be aware of what learners need to properly enhance learning through assessments.



**Figure 3 Assessment Pyramid**

The middle-level of the pyramid is *Assessment for learning* which occurs when teachers use assessment to identify, track and measure students' progressions. An important factor of this type of assessment is the contribution from the teachers to the learners throughout effective feedback and appropriate guidance. Earl (2003) points out that teachers need to plan strategic assessment to measure what the learners have acquired and to which extent they are skilled in several language practices. Teachers need to use a range of tools and techniques to explore the competencies and performances of the language learners. Bell and Cowie (2001) define this type of assessment as the process of working collaboratively with teachers and learners to improve their methods to provide a better quality of learning and teaching in the classroom. Finally, at the bottom of the pyramids is *Assessment of learning* which occurs when teacher collects learning evidence to critically evaluate the whole progress of the learner and teachers make judgments whether the learners should or should not upgrade to the next learning stage, in an educational context. These assessments have direct impact on the students' decisions since the outcome would demonstrate if the learning goals and needs were met.

This pyramid allows to visualise assessment in three perspectives, taking into consideration three types of evaluation: the learner, the teacher-learner and institution. In each of these three layers, the progression of the student is presented for the purpose of completing an accurate evaluation process. Students' progress is defined by Masters and Forsters as "*a description of skills understanding and knowledge in a sequence in which they typically develop*" (1997:1). Institutions tend to have different parameters of the use of assessments. Some institutions can be rigid by taking only the test into account for the purpose of meeting a specific aim. On the other hand, some institutions could include more aspects such as class participation or homework. Either scenario students are assessed based on aspects divided into blocks of knowledge to be demonstrated by different types of assessments. Popham (2007) indicates that progress is a selected criterion/content for a specific domain that enables knowledge, techniques and skills. Hence, we can see progression as the acquisition of different skills, techniques and practices that improves cognition of certain content or domain.

It is worth mentioning that language learning is a discipline that does not share qualities with other domains in which teachers can capture progression easily. The use of language can be seen for two main contexts: the social interaction and the educational interaction. Research studies have indicated that learners tend to increase or have a positive progression when language can be used as a social component (Masters and Forsters, 1997). This positive progression does not indicate that the level of proficiency is higher, but the level of confidence can increase since they are not being judged by any member of a society. In an educational interaction, especially in a language classroom, teachers are constantly indicating what learners need to acquire and put into practice with the intention to increase the proficiency level.

Language teachers should be aware of techniques and skills on how to measure and track students' progress based on students' performances (tests' scores), which is the time of instruction between each assessment. On one hand, learning is a process in which the teacher is the enabler of knowledge and s/he requires to plan and implement different teaching practices to be able to deliver the knowledge to students. Performing, on the other hand, is the process in which the classroom teacher administers a variety of assessment tools to observe and measure how students' knowledge has been retained and tested. In an educational utopia, students are expected to make positive progress from the beginning until the end of a language course; however, research studies have shown that students' progress can fluctuate with peaks and downs in their learning journeys. It is also worth indicating when tests' scores do not always show improvement, as this does not mean that the student requires more instruction; in fact, this shows in where the students are 'stuck up' with a particular language issue and requires more

attention to tackle this. Gass and Selinker (2001) refer to this as “learning plateau” that indicates the student has not progressed enough to reach the learning goals from a course.

Knowledge of language, knowledge of the purpose and reasons for assessment and knowledge of the different aspects of assessment allow teachers to select and adapt resources for their teaching practice to maximise learning. Nevertheless, the institution where students and teachers are constantly involve in assessment might have different perspectives on how to implement testing materials. That is the reason why knowledge of the institutional educational polices is fundamental to improve the teaching practice.

### **2.4.1.3 Institutional Assessment Policies**

The foundations of knowledge of the language and how the language is being taught and learnt are example of what teachers are assumed to identify and implement in a language classroom. However, there are other factors that influence teachers’ assessment literacy. Literature has suggested that teachers’ assessment skills and practices are influenced by the teachers’ backgrounds in relation to their experience of teaching from their institutions as learners and now as teachers involving assessments. Practitioners’ trajectories are sometimes associated to how the teacher understand and reflects on the institution where the language is learnt and taught. These aspects may influence the assessment practices and unable teachers to develop a solid knowledge. The institutional contexts in which teachers and programme managers work play an important role to understand how they operate assessments.

Assessment has a direct impact on the social and cultural perspectives and the relevance of the educational context. Inbar-Lourie (2008b) indicates that there is a lack of understanding between two different types of national perceptions when assessment is carried out. The testing and assessment cultures are a factor that drives assessment with a social interaction and educational considerations. The testing culture is the psychometric view on how tests are constructed, validated and used. Some countries believe that that assessment holds an important aspect of educational purposes throughout the administration of tests. On the other hand, there is the assessment culture, which requires understanding on how learners and teachers interact through different assessment tools and methods. The assessment culture requires comprehension on how learners acquire the vital knowledge through their social and cognitive processes. Moreover, there other relevant features that are linked with the previous set of cultures: the historical evaluation, the philosophy behind the use of assessments and the responsibilities from the ministries of education to empower the educational assessment system in a country.

It is under the syllabus and curriculum in which the institution relies on the methods for assessments. The educational institution plays a fundamental role in the process of making decisions on how students and teachers are involved in the act of assessing the language. The two previous stakeholders and the programme managers require to understand the methods in which they tend to evaluate and assess language proficiency. However, the institution is not the only body that regulates this process, in fact, the local and national ministries of education have a significant impact on how to apply, interpret and reflect on the assessment practices (Lynch, 1996; Inbar-Lourie, 2008a). It is vital teachers understand the cultures of testing and assessment since this difference could allow stakeholders comprehend the assessment purposes and processes.

Fulcher (2012) highlighted that one of three reasons for assessments goes beyond the place of learning, the classroom, but the impact of tests and results have a significant influence on society (Shohamy, 2008). Institutions, from time to time, administer national and local assessment that enable to provide beneficial attributions towards the assessment processes among students, teachers and other stakeholders such as directors, programme managers or government (policymakers). Identifying assessment strategies with certain purposes allow stakeholders to interact more adequately whilst instructing the language, as well as comparing students' performances. Understanding and practicing the assessment methods in which stakeholders are familiarised could empower the decisions of selection of language testing and assessing materials to benefit institutional policies. However, the misuse or the lack of skills in identifying the rationale behind the proposed national policies could influence the process of learning and scores. Another aspect that institutions require is the comprehension on how assessment and evaluation could assist teachers, students and other authorities in identifying areas of improvement; not only for the instruction, but other areas that work in collaboration to evaluate the programmes and institutions. In Section 2.5 the situation in relation to the Mexican educational system and institutional assessment operational systems are presented in detail.

Teachers who have worked in their institutions for a significant period of time might be unaware of the current situation in terms of how the government has influenced educational decisions (SEP, 2010; Popham, 2011). The national measures play an important role and allow institutions and government to take decisions on how to improve the current situation following the national standards. The following table (see Table 3) presents the matrix of the skills and practices related to the three standards of the component of knowledge of assessment.

**Table 3 Matrix of Knowledge of Assessment**

Component	Standard	Skills and Practices
Knowledge of Assessment	Applied Linguistics	<ul style="list-style-type: none"> <li>a) Understand language use and application into the classroom</li> <li>b) Application of methods for learning, teaching and assessing (memorisation, drills, task-based learning)</li> <li>c) Theoretical knowledge and application of learning trends (behaviourism, constructivism and socio-cultural)</li> <li>d) Justify Second language theories affective approaches (motivation, preparation, learning efficiency methods)</li> </ul>
	Educational Assessment	<ul style="list-style-type: none"> <li>a) Awareness of assessment for/of/as learning</li> <li>b) Understand the underpinning of content and purpose of assessments</li> <li>c) Plan and implement materials for assessing language in relation to the learning context</li> <li>d) Enhance and maximise learning through assessment practices</li> </ul>
	Institutional Assessment Policies	<ul style="list-style-type: none"> <li>a) Understand of the socio-political aspects of assessment towards national policies</li> <li>b) Apply national and institutional standards into the language classroom</li> <li>c) Distinguish assessment methods that benefit individuals, institutions and educational sectors</li> <li>d) Participate in development of tests and assessments for classrooms purposes</li> </ul>

This sub-section, knowledge of assessment aimed at reflecting on whether teachers are prepared with suitable knowledge of the methods for learning and teaching the language that they are assessing. Understanding the culture and the foundations of the sociocultural integration in the classroom setting would allow teachers to identify methods for improving progression. That is why is vital that teachers be aware whether they possess solid knowledge of applied linguistics, educational theories and knowledge on how assessment should be treated at local and national scales. The following section introduces the second component of the teachers' language assessment skills and practices, technical considerations in assessment.

#### **2.4.2 Technical Considerations in Assessments**

This section includes four educational assessment standards: 1) testing principles, 2) assessment purposes, 3) assessment processes, and 4) educational measurement. These four standards are in relation to technical considerations of assessment and teachers that require understanding and knowledge to maximise teachers' assessment skills and practice in the language classroom.

### 2.4.2.1 Testing Principles

Different stakeholders (e.g., test developers, programme managers and language teachers) are required to develop, implement and adapt testing materials which meet the learning objectives of a course. Developing language assessment instruments could be considered a complex process due to quality assurance implications in which they are prerequisite to be validated and reliable (Harrison, 1983; McNamara and Roever, 2006). In order to construct language testing instruments, these stakeholders, need to follow specific principles or foundations to successfully administer, measure and mark accurately to avoid any bias for learnings and also institutional policies. In the book of '*Language Assessment: Principles and classroom practices*' by Brown and Abeywickrama (2010) highlighted five important principles with regards to test development: validity, reliability, practicality, authenticity and impact (washback). In addition, these principles are also recognised as '*usefulness of a test*' (Bachman and Palmer, 1996: 18). Table 4 defines each of the five previous mentioned principles and their definitions, as well as their applications.

**Table 4 Testing Principles: Definitions and Applications**

Testing Principles	Definition and Applications
Validity	The extent to make sure the instrument has a clear purpose, and it actually assesses what it supposes to test.
Reliability	Focuses on the test consistency in specific group of test takers. The content and the results are reliable when the scores are evaluated to identify consistency. Also, students should obtained similar scores if the test is designed under the same linguistic ability.
Practicality	Deals with resources, methods for scoring and the time management to develop and complete a test for its final administration. Institutions should be aware of the facilities and resources they have so as to administer tests and other assessments.
Authenticity	Focuses on the use of language in certain real-life situations in which test takers can apply certain knowledge outside the classroom.
Impact (washback)	The effect from the test scores and how much impact had these towards the institution and society.

Bachman and Palmer used the term *usefulness* to identify and examine the properties and characteristics of testing instruments. These five principles have been explored in different research studies in relation to language assessment literacy. Tsagari and Vogt (2017) considered testing principles as one of the areas language teachers in Europe lack of training. Giraldo (2018) also included testing principles in his core list of language assessment literacy elements that allow teachers to implement better resources to test what students have previously learnt. Validity and reliability are, to some extent, the most commonly discussed terms in the literature due to the importance with regards to the development and design of tests. However, it is relevant to not avoid the other principles (i.e., authenticity, practicality and impact) whilst developing the blueprint of language tests.

## Chapter 2

To begin with, validity focuses on the structure and fitness of the items and tasks within a test that need to measure what they are intended for (Fulcher and Davidson, 2007). Messick (1991) defines validity as the overall evaluation of a test to scrutinise that the main purpose of a test is being conducted with a specific group of stakeholders. Similarly, Kunnan (1998) defines validity as a property of classroom language tests in which this is linked to what the blueprint of the test should reflect on the curriculum aims. This principle lingers with the content, construct, criteria and structure of a test and its items. It measures the intentions of content based on the instructional practice. There are several types of validity, but the most common types for language tests are the following:

- Content validity- this type of validity scrutinises aspects related to the contents, knowledge and skills of what the test is supposed to test. This is to identify if the tests contents were achievable by the test takers.
- Construct validity- this type of validity focuses on the way items and tasks were developed for tests to meet constructs. Brown and Abeywickrama gave examples of constructs: “*proficiency, communicative competence and fluency*” (2010: 33) If a task that is supposed to assess communicative strategies, items and tasks should reflect the contents and the construct and how the items would benefit the learners.
- Criterion validity- this type of validity infers that the test is following a certain type of criterion that allows teachers and students to position their performances in a measurable scale. Therefore, this validity should reflect the skills and knowledge of a test taker into a scale. Brown and Abeywickrama claimed that there are two more types of validity under the umbrella of the criterion validity. The concurrent validity is which test scores are compared with other scale or performance. The predictive validity is to assess and examine the possibility of test takers upgrading to the next learning stage.
- Face validity- this type deals with the appearance of the tests and how they look in terms of their structure and contents. Tests should follow certain institutional guidelines to make it adequate for its administration. Researchers have identified that this could not be considered a type of validity. However, classroom teachers tend to use this as a method of practice (Hickey, 2014).
- Consequential Validity- this type of validity deals with the test effect on the socio-political aspect (macro-level) and effect on individual test-takers (micro-level) (Bachman, 1996). This type of validity is used for high-stake tests, but teachers should possess certain skills and practices to identify the impact of a test inside and outside the language classroom.

According to Weir (2005) who developed a recent perspective to combine validity into cognition, socio-cognitive approaches and the use of tests with the plausible impact from the evaluation. He developed a framework entitled as socio-cognitive framework. His contributions

towards validity with the inclusion of socio-cognitive approaches made this a unique source of possibility to use validity in other approaches.

Having mentioned the importance of having adequate skills and practices to understand the principle of validity, it is also essential to distinguish why reliability plays an important role in the usefulness of tests. Reliability refers to the condition in which tests can be administered twice or more to demonstrate irregularities in the test-takers' scores. Bachman (1990) and Bachman and Palmer (1996) describe reliability as the consistency of the measurement and criteria that was applied to measure a test-takers' performance. The development and use of marking schemes (marking criteria) needs to follow a well-constructed sequence that can facilitate its use and application. A misuse of marking criteria could be the lack of knowledge, skills and practices in which examiners might disagree in test-takers' performances. There are four types of reliability that can be conducted into tests:

- Student-related reliability- this type of reliability relies on the psychological and external factors of the learner. If a student presents irregularities such as illness, anxiety or a language disability, this could affect the score of the test taker in their performances.
- Test administration reliability- this type of reliability deals with the institutional procedures of administering language tests. Location and quality of the testing materials might influence the test-takers' scores by affecting their performances.
- Rater reliability- this type of reliability deals with the people, mostly language teachers, who assess test-takers performances in the language classroom. Teachers with good assessment literacy levels could have more skills and practices to assess students; whereas a teacher who has poor skills and lack of practices in assessing students could affect students' performances. There are two types: inter-rater reliability is presented when two or more scores are equal or alike in consistency based on the criteria being used. On the other hand, intra-rater reliability when the rater consistently marks tests and assessment practices, but these can be affected due to external factors that can affect students' performances.
- Test reliability- this type of reliability deals with the structure of the test in terms of items and tasks quality. If an item is poorly written, this could lead to a confusion from the test takers. Revision of items and tasks should be taken into consideration.

Teachers with high levels of language assessment literacy should understand the differences and the purposes in which these two principles, validity and reliability, are applied into the planning, developing and administering stages of language testing instruments (Popham, 2006; Stiggins, 1991, 1995, 2002). The effectiveness of the testing instruments could be greater when assessment literates are able to distinguish between these two principles and their sub-

categories. It is important to mention that practicality, authenticity and impact (washback) are also crucial for the development of instruments and teachers should consider applying them into their language classrooms, but more essentially in the institution (Bachman, 1990, 1996).

Although these above-mentioned principles are not frequently discussed in literature in language assessment literacy for language classroom teachers, their relevance should be highlighted and not avoided (Brown, 2012). Teachers with good assessment literacy levels can differentiate if a test can meet the criteria in terms of its practicality taking into consideration relevant aspects; for instance funding, time constraints, methods for its administration, human and material resources, as well as ways of marking. Also, teachers should be able to measure its authenticity and impact of tests. The authenticity of a test comes with certain qualities such as how authentic a test could be, the structure of items and tasks should be contextualised, meaningful and relevant. Lastly, the impact of school-based language assessments might not be as relevant as those tests such as high-stakes, teachers should be skilful in identifying which sources would have more impact on the learner as well as for teaching, learning and assessing language practices.

These testing principles have been explored to identify teachers' language assessment literacy in terms of developing language tests, however, there has not been research conducted in Mexican classrooms that indicates whether the teachers have any skills and practices in relation to the use of testing principles. Even if language teachers are not part of the process of developing language tests, they should be aware of applying these principles to improve the quality of the testing resources administered to their learners which aimed to be accurate, validated, reliable and meaningful for their language learning trajectories. The following section introduces the standard of assessment purposes. This standard presents the skills and practices teachers need for empowering language assessment literacy.

### **2.4.2.2 Assessment Purposes**

Language teachers need to be aware of the purposes and reasons why tests and assessments are administered in the classroom. Stakeholders should have clear objectives to choose adequate assessment methods and administer adequate testing resources. Developing skills and practices could essentially bring a significant number of benefits to understand how students tend to perform, as well as for teachers to revise and identify irregularities in their teaching and assessing practices. Teachers are expected to examine and track' students' progression by exploring how much knowledge and how they have acquired this in a specific period (Bachman, 1996).

Cohen (1994) listed three main purposes for using assessments that could be understood based on stakeholders' intentions. The programme managers, at their institutions, might require assessments to identify different aspects of the curriculum and institutional management among teachers. Cohen indicated that administrative assessments, those which deal with the placement of the students, certification and promotion to higher levels within an institution, are vital to accommodate students in the right language course. The second purpose is instructional in which teachers should be able to administer and highlight how students have learnt. Cohen describes this as diagnosis, evidence for progress and a proper evaluation for the curriculum. Lastly, the last purpose was for conducting research to examine and evaluate the efficiency of language learners in a course. Teachers should be aware of the skills to develop, adapt and use a test for the objective of its administration.

Nevertheless, Cohen's rationale for these assessment purposes does not always relate to what teachers currently require for administering language tests in the classroom. Brown (2012) indicated that the instructional purposes focus on those assessments that are for classroom purposes with the intention to measure test-takers' progress and achievement. This researcher also highlighted that teachers tend to be mainly interested in identifying how learners have progressed in a particular course. According to Brown, there are three purposes for classroom-based language assessment. There are diagnostic, progress and achievement. Table 5 describes each of these purposes. This table also provide the definition, application and type of assessment based on Popham's (2011) taxonomy of assessments.

**Table 5 Purposes of Classroom-based Assessment and Applications (Popham, 2011)**

Purpose of Assessment	Definition	Application	Type of Assessment
Diagnostic	They provide a holistic evaluation of what students have learnt and what they need	This type of assessment can be administrated at the beginning of a particular course to evaluate what learners have acquired in previous courses.	Formative
Progress	They indicate how much students have acquired in specific period of time	Teachers administrate these during the course to evaluate what the students have acquired in a period of instruction.	Formative
Achievement	They focus on how students have performed in the course and teachers make judgments to promote the students to the next level	This assessment is administrated at the end of the course	Summative

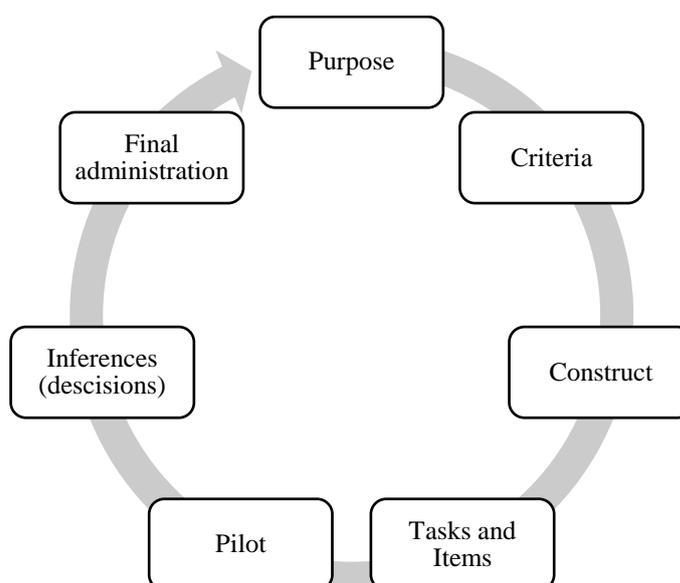
Institutions in Tamaulipas, as any other Mexican state, use these three purposes for language assessments at different stages of a particular course (SEP, 2012). However, each institution has the flexibility to adapt the type of assessment based on their institutional needs. Teachers require to differentiate the purpose of the assessment when tests and assessments are

administered with the intention to select the suitable tasks and teachers could expect appropriate responses (Rea-Dickins & Germaine, 1992). A number of educational institutions have certain restrictions in terms of the development of the achievement tests (summative assessment) which affects the institutional operational assessment system. Programme managers are sometimes responsible for developing summative tests with the intention of standardising the methods of teaching, learning and assessing language in each institution. Language teachers occasionally can assist to develop and revised the language tests but this activity relies more on less objective face and content validity. Interestingly, the international and national market of language testing has significantly increased in the last decade and currently there is a large number of course-books equipped with repertoire of tests created by certified publishing houses (Malone, 2012; Fulcher, 2012).

Even though some teachers are not able to participate in the development of language tests for their institutions, they need to be aware of the differences in purposes for language assessing materials. By distinguishing the purpose of assessment, the types of items and tasks, different language assessment methods, teachers and LPMs could increase their language assessment literacy perceptions. Appropriate skills and practices are important as teachers can actually enhance learning by developing quality assessment materials. The following section discusses the standard of assessment processes.

### **2.4.2.3 Assessment Processes**

The development of language tests and school-based assessments can be described as a tedious process which takes different elements into account (Bachman, 1990; Bachman and Palmer, 1996; Hughes, 2013). Teachers should understand the different stages of the development of language tests, not solely for standardised tests but also for the classroom and institutions (Brown, 2012). Fulcher and Davidson (2007) developed a framework (see Figure 4) that illustrates the assessment process. They considered seven stages that take part of the process of developing language tests: a) test purpose, b) test criterion, c) construct definition, d) task specifications, e) pre-pilot the instrument, f) making decisions, and g) final administration.



**Figure 4 Fulcher and Davidson's (2010) Test Development Framework**

In order to develop appropriate and effective tests, it is important to understand the purpose and use of such instruments. Test purposes are aligned to test-takers' needs, the contents from the instruction and possible items which students are familiarised with. Henning's (1987) taxonomy of tests uses and purposes can be applied, not only for the linguistic performance, but also for measuring other academic disciplines.

- ***Objective tests:*** provide clear arguments why such items have only one or a specific reasonable response. The use of a test key might be required to mark the items accordingly to an argumentative rationale.
- ***Subjective tests:*** deal with open responses and it is very likely that this method of marking the items of this particular test provides more than one possible answer.
- ***Direct tests:*** use for measuring specific testing situations in which real-context language is part of the method of marking the item. For the skill of writing, drafting a couple of sentences with regard to an authentic trend topic (e.g., global warming, health and fitness, etc.)
- ***Indirect tests:*** assess different techniques in which real-life language is required but context needs to be objective to be specified assesses these tests.
- ***Discrete-point tests:*** assess one specific language skill (e.g., reading, etc.) or language component such as grammar, vocabulary or pronunciation.
- ***Integrative tests:*** integrate a variety of language abilities that the performance can attempt multiple tasks at once.
- ***Criterion-referenced tests:*** these are assessed by holistic and analytic criteria and do not required other test-takers' performances.
- ***Norm-referenced test:*** require other test-takers' performances to measure the score based on the group's performances.
- ***Ipsative tests:*** measure the progression of the test-takers by comparing previous performances and it also focuses on test-takers hours of instruction and their own previous performances.
- ***Speed tests:*** these tests are assessed accordingly with the pace of test takers needs. These tests mainly assess the development of knowledge and speed time.
- ***Power tests:*** design to be answered with specify period of time. Test takers should demonstrate they are capable to develop the knowledge accordingly to time speed.

## Chapter 2

Criteria and construct are in relation to the task specifications. These are the underpinning that states the requirements for the items and the contents with the aim of developing the instruments to assess (Fulcher, 2012). In institutions, criterion covers aspects such as the total number of points available for a task. In this case, teachers take into consideration the possible points for each item, task or even the test. In terms of the contents, teachers should be using the content that they were delivering during the class since students need to be tested on these.

The use of task specifications is normally applied with those high-stake tests that require to document each of the items and topics included in a test. This document, task specifications, varies depending on the type of tests, skills tested and other essential aspects from the test takers. In language testing, it is important to mention the linguistic competence that is being assessed and the way in which test takers perform are sometimes misled. Some features teachers should take into consideration at the moment of developing language tests are the following:

- Type of task (e.g., Multiple-choice exercise, verifying information)
- Task level and difficulty (e.g., CEFR: A2, B2)
- Number of items
- Test-takers' personal characteristics (defining the construct)
- Methods and score procedure (e.g., 1 mark per item)
- Length of the task (time)
- Cognitive and linguistic processes
- Language skill or linguistic component (e.g., Speaking/ Grammar)
- The method of delivery (written or aural)
- The environment of the testing situation (paper-based or computer-based tests)

The task specification needs to be revised several times by a group of testers or test developers to verify that the test actually aims to assess the required content. With regards to the development of items, it is important to have in mind the objective, the test-takers' background, criterion for the items and the methods of testing. For teachers who develop their own tests, they should have certain familiarity with the aforementioned features to make the process of assessment accurate and authentic. In addition, choosing the right items to develop a task or a test could bring negative outcomes if teachers lack of knowledge for selecting appropriate items in particular assessments. Some types of items frequently seen in language tests, not solely in the classroom but also standardised tests are the following:

- Verifying information (e.g., true, false, does not say)
- Matching items
- Classification items
- Multiple-choice items
- Short-answer items
- Fill-in-the-blank

- Sentence completion
- Cloze items
- Extended response

Piloting tests is the first stage of identifying what has worked well and what needs to be amended at the movement of developing tests. Once the piloting stage has taken place, teachers can identify and modify any problems, if any, related to the structure and content of the tests. The final administration of a test can be through paper-based, computer-based, or in the case of speaking, this could take form of a presentation or a roleplay. Once all test-takers' (students') results are gathered they could also be evaluated through item-based and test-based levels to measure their efficiency towards the purpose of the test. Statistical analysis (see Section AII) can provide a large number of advantages in deciding which items or tasks may be recycled in order to develop an item bank for future testing situations.

Stakeholders with assessment literacy should be aware of the different types of items and tasks in relation to what teachers should assess language regardless their working contexts (Stiggins, 1999, 2002; Brown, 2012). It is worth noting that not all teachers require to be aware of the development of school-based assessment since the context in which they work, but considering that these are important considerations at the moment of developing language tests, programme managers, teachers and other stakeholders require to be attentive and understand these steps. In terms of the assessment practices, understanding and applying these processes can bring a large number of advantages to the development of testing instruments and teachers' professional development. The following section discusses the evaluation procedures of items and tests to maximise its efficiency and meet the purpose for using these educational tools.

#### **2.4.2.4 Educational Measurement**

When language teachers listen to the word 'measurement' they tend to relate this to numbers or statistics, which could be a negative aspect for educators. Literature suggests that teachers avoid statistical analysis with their language tests due to the lack of time, fear or simply because they do not have enough skill on how to conduct this type of analysis appropriately (Malone, 2008; Fulcher, 2012). Test developers and testers tend to use statistical approaches with the intention to make their items and tasks much more valid and reliable. These approaches are mainly used for standardised testing since these instruments require a high degree of reliability due to high-stake purposes such as accountability and mobility. These approaches assist to identify those items that might affect the performance of the test takers, as several items could fail to assess what they were supposed to.

Technically speaking, there are two main trends for conducting this type of analysis. Classical Test Theory (CTT) and Item Response Theory (IRT) are the theories pertaining to item and

test analysis. These theories offer a number of evaluations that could help teachers create better items with the intention to re-use them for further testing situations (Bachman, 2004). There are different methods for analysing items in language tests and nowadays teachers have easy access to these types of statistical analyses. The *index of facility* and the *index of discrimination* are the most common methods discussed in the literature for language teachers: the first one evaluates the difficulty and the second the discrimination between with higher or lower scores (Hughes, 2013; McNamara, 2004). I have chosen these two statistical functions for this research since these two can be conducted for language classrooms.

According to Alderson, et al. (1995) the Item facility index measures the level of difficulty of the items by calculating the average of the correct responses with the total number of test takers. If the p-value ( $p$ ), statistical term for this method, is close to  $p=1.00$ , it means the item is relatively easy; whereas it is close to  $p=0.00$  means the item is relatively difficult. On the other hand, item discrimination index measures the quality of the items by dividing the group of test takers into higher and lower scores. This statistical test is known as point-biserial correlation ( $rpb$ ). If the correlation is close to  $rpb= 0.00$  or negative, the items need to be revised or discarded from the test, whereas if the correlation is above  $rpb= 0.20$ , the quality of this item is satisfactory in the test. Another method to analyse test items is the functional distractor analysis. This type of analysis is conducted by scrutinising the properties of multiple-choice items by obtaining the p-values in each of the distractors. Literature suggests that those distractors which get lower than 5% of the responses or higher as 95% should be discarded because of the lack of responses or similarities with the key option.

Language assessment literacy deals with the skills and practices to conduct these procedures. A language assessment literate should be of recognising how to calculate or at least interpret these indices to evaluate test items. Arguably, not all teachers could measure these indices, but language programme managers may take into consideration these procedures as part of the evaluation of the testing resources as an institutional policy. The knowledge of educational measurement is essential due to the large number of advantages that may provide to improve not only the quality assurance for school-based materials, but also to improve learning and teaching materials and practices. At institutional level, language programme managers are desired to deliver high quality tests to their teachers; therefore, they are expected to conduct this type of item analysis in those institutions where assessment could be slightly restricted to all language teachers. However, teacher could also apply these two techniques to identify if the materials such as worksheets or quizzes were able to meet their purposes. Table 6 illustrates the TLASP for the component of technical considerations in assessment.

**Table 6 Matrix for Technical Consideration in Assessments**

Component	Standard	Skills and Practices
Technical Considerations in Assessments	Testing principles	<ul style="list-style-type: none"> <li>a) Identify the properties and usefulness of a language test by applying the testing principles.</li> <li>b) Demonstrate awareness of the relevance of the principles at the moment of developing language test.</li> <li>c) Classify language tests based on their usefulness and application towards learning.</li> <li>d) Judge the usefulness of language tests towards teaching and learning practices.</li> <li>e) Apply knowledge to highlight irregularities and efficiency in language tests.</li> </ul>
	Assessment purposes	<ul style="list-style-type: none"> <li>e) Understand the differences of assessment purposes</li> <li>f) Apply accurate assessment method such as diagnostic, formative or summative</li> <li>g) Choosing the purpose of tests based on curricula and learning needs.</li> <li>h) Contrast and examine differences among types of assessments</li> <li>i) Examine the different assessment purposes to match the teaching and learning objectives</li> </ul>
	Assessment processes	<ul style="list-style-type: none"> <li>e) Understand different properties and usages of language tests.</li> <li>f) Differentiate and choose appropriate criteria.</li> <li>g) Underpin the purpose with the method for assessing a particular language skill or linguistic component.</li> <li>h) Awareness and application of different types of items and tasks to be included in language assessments.</li> <li>i) Infer and make educational judgments about the qualities of test.</li> </ul>
	Educational measurement	<ul style="list-style-type: none"> <li>a) Awareness of methods for evaluation of items and tasks</li> <li>b) Identify the level of difficulty of items and tasks</li> <li>c) Differentiate between higher and lower scores to discriminate items</li> <li>d) Conduct analysis to measure the quality of items and tests</li> <li>e) Distinguish functional and non-functional items and tasks</li> </ul>

It is important to highlight that some teachers develop their tests without taking educational measurement into consideration. Research has suggested that these statistical tests could bring a large number of benefits to the tests for high-stakes tests (Cohen, 1994). However, it is also essential to identify if some of these statistical tests were applied in the language classrooms; especially those tests that require to be administered more than once.

### 2.4.3 School-Based Language Assessments

Under the umbrella of school-based assessments, four educational standards have been addressed: 1) language assessment, 2) assessment for learning, 3) communication of assessments and results, and 4) assessment as an educational support.

### 2.4.3.1 Language Assessment

This section focuses on what teachers require about the methods for assessing language skills: the receptive and productive. The researcher considered to include this section into this component because of the frequency in which teachers deal with language assessment and choosing the appropriate types of items and tasks for a particular skill. It is indispensable teachers acquire certain skills and practices to assess productive and receptive skills separately because they assess different approaches, strategies and linguistic features (Fulcher, 2003; Luoma, 2004). The assessment of productive skills, speaking and writing, is mainly conducted by using marking schemes, criteria or rubrics. Teachers would use appropriate marking criteria that match with the type of assessment, language contents, and the test-taker's language level (Brown and Abeywickrama, 2010). This process is likely to be less objective because test-takers' responses are collected and then assessed by examiners or teachers who might lack of skills to conduct this adequately. Conversely, receptive skills, reading and listening, are more likely to be objective since they rely on how the items were constructed.

The marking criteria for productive skills need to be chosen carefully with the intention to have a fair and less bias assessment. Language teachers need to take into consideration several factors such as: students' level, the contents, the type of task and what linguistic features are the ones to be assessed. There are two types of marking criteria for productive skills: the holistic and analytical. The holistic provides an overall perspective on how the student has performed in a task. Nevertheless, the analytical assessment takes into consideration a variety of linguistic features depending on the language skills (Rea-Dickins and Germaine, 1992). The features for assessing speaking could be grammatical competence, lexicon competence, comprehension, fluency, pronunciation and the coherence. In the assessment of writing, on the other hand, assess the organisation of the ideas and the content, grammar, punctuation, spelling, mechanics and the written style.

The receptive skills cannot be assessed with marking criteria; they are in fact assessed based on items previously revised and validated in order to assess certain techniques or skills that reading and listening comprehension require (Tarone, 2005). It is worth mentioning that these skills could be assessed with mark schemes only if the productive skills are being integrated in a task. In terms of reading skills, Han and Anderson (2009), suggest 5 different reading strategies for assessing the structure of a passage. First, the understanding of the whole passage in which students need to be aware of the type of audience, sequence of events, distinguish facts from opinions and recognise the structure and genre. The second is at the context level, in which test takers need to be aware of the subheadings and its logical organisation. The third strategy is at paragraph level, in which test takers need to distinguish between statements and examples,

understand the main and secondary ideas, and finally infer the content. The fourth and fifth strategies are across at sentence level; here the test takers need to identify the functions of the words (e.g., nouns, verbs, etc.) to complete certain items or tasks. These are several assessing strategies that teachers would be made aware to promote and practice in the language classroom with the intention to increase students' scores.

With regards to assessing listening skills, Buck (2001) indicated that there are three different approaches to assess listening. The first one is the discreet-point approach, which focuses on understanding isolated words to be tested in simple tasks. This approach is to distinguish phonemes, recognise words from a monologue or a dialogue, however, it has been criticised for the last decade as an obsolete method with regards to the type of performance. The second approach is the integrative in which test takers need to understand a dialogue and grasp words that are suitable for the type of conversation. The last approach is communicative; test takers are required to listen to the recording in order to understand and clarify what the situation is, infer the context and register of the conversation with a personal-spoken response.

Items and tasks, for both receptive and productive skills, are necessary to be understood by teachers and test developers since some items are more suitable with certain tasks and language skills. Brown (2012) indicates that item responses can be intentionally linked with the purpose of the assessment. This researcher highlighted three main types of responses, selected, productive and personal. Table 7 illustrates Brown's (2012:135) three types of responses and their applications for the language classroom.

**Table 7 Types of Responses and their Applications (Brown, 2012: 135)**

Response	What to assess?	Item categories	Application in the classroom
Selected	Linguistic properties (e.g., grammar and vocabulary), receptive skills of reading and listening	True or false Multiple choice Matching Confirming information	Exercises with grammar, vocabulary punctuation and isolated information for reading and listening tasks
Productive	Productive skills (speaking and writing) or integrated skills	Fill-in the gaps Short answer Communicative task Cloze items	Productive knowledge, pair work, role plays, etc.
Personal	All four skills, degree of understanding of certain topic	Self-/peer assessments Portfolios Conferences Interviews	Reflection about a task or specific language content, project work, individualise instruction, etc.

It is worth mentioning that some teachers who are not part of the development process of tests and testing materials in their institutions may find it difficult to appropriately use the marking schemes for speaking and writing. Literature suggests that some teachers assess with the same methods as they were assessed (Mertler, 2002; Newfields, 2006; O'Loughlin,

2013). Therefore, teachers need to be prepared with specific knowledge that enables to have a better understanding of the applications of the scoring procedures. By understanding the differences between holistic and analytic marking schemes, language teachers and programme managers could contribute to an improved assessment practice in their institution. These skills required for teachers so as to provide learners with enriched and appropriate opportunities for a fair and less biased assessment. Teachers with a higher level of language assessment literacy can use different assessing strategies, methods and skills with the intention to increase the quality of the materials and improve the teaching and assessing practices (Stiggins, 2002). The following section discusses the standard of assessment for learning which deals with those alternatives, not language tests that teachers can use to assess their students and to promote learning.

### **2.4.3.2 Assessment for Learning**

Assessment for learning (AfL) plays an important role in the language classroom-based assessment not solely for the language learners but also for the teachers. This term mainly relies on the formative assessment, which deals with the students' progression and the use of different assessment tools in a particular course. AfL assists teachers to discover their students' learning trajectories and how learners have gained linguistic knowledge. Araujo (2010) and Martinez Rio (2013) highlighted that there is a lack of formative assessment practices, also known as assessment for learning, in the curricula. Teachers tend use traditional assessment approaches such as the final product (final test) to upgrade to other learning stages. Teachers should be aware of the different assessment tools and their applications to empower the learning, teaching and assessing methods.

Teachers should require a deep understanding of the different assessment tools in order to use them to successfully observe the progress of the students. Implementing (AfL) activities in the language classroom may enhance teachers' language assessment literacy development. Hamp-Lyons (1997) listed the characteristics between performance assessment and standardised test. Table 8 presents some of characteristics of these types of assessment are presented.

Language classroom teachers need to differentiate between a standardised test and an instrument for testing in the language classroom. Up to here, the only tool that has been discussed for measuring somebody's linguistic competence and performance in this report has been a test. Nevertheless, there are different assessment tools that teachers can apply in their lessons to have a better perspective of what assessment is and how to foster students' learning autonomy.

**Table 8 Performance Assessments vs Standardised Tests (Hamp-Lyons, 1997)**

<b>Performance Assessment</b>	<b>Standardise tests</b>
Criterion referenced	Norm referenced
Contextual objectives	Decontextualized objectives
Modifiable	Uniform
Multidimensional	Restricted dimensions
Longitudinal	Pre/post administration
Continuous recording	Discontinuous recording
Monitors progress	Statistical view of achievement
Extensive behaviour sampling	Restricted behaviour sampling
Reflects quality of work	Reflects speed and accuracy
Promotes student learning	Promotes skill in test-taking
Enhance student motivation	Promotes student anxiety
Instructionally relevant	Instructionally independent
Contributes to classroom change	Imposes institutional change
Informs instructional decisions	Justifies bureaucratic decisions
Useful to parents and others	Unhelpful to parents and other

The materials for learning a language vary depending on the curricula and the purpose of the course. University learners might not be that interested in colouring or playing some games like young learners. Brown and Abeywickrama (2010) highlighted that performance-based assessment are activities that measure the progress of students' performances. These are not standardised methods of assessing language, but they convey several advantages to students in identifying what they have learnt and what they need to learn. Table 9 lists several activities and their applications in higher education.

Institutions in Tamaulipas implement a variety of activities that focus on the learners' progress. However, the most important aspect is the summative assessment. According to the needs analysis, teachers use language portfolios in both private and public institutions. The use of portfolios can assist teachers and learners to track their learning progress, but at the same time, teachers can also identify whether certain materials are suitable for their teaching practice (Genesee & Upshur, 1996). It is worth stating that a test is the main product that higher education language teachers in Tamaulipas look to make any judgments before upgrading students to higher learning levels.

**Table 9 Other Assessment Tools and Their Applications**

Assessment tool	Applications	Type of Assessment
Portfolios	It focuses on the development of the learners' learning progress. It collects relevant learning evidences that have a direct impact on the purpose of the course.	Formative/Summative
Rubrics	These are not types of assessments but tools that assist teachers and learners in identifying what they need to learn and what they already know in terms of language	Formative
Journals	Learners can use journals to describe their days, learning experiences and what they want to achieve academically or personally. This is a tool in which students are free to write but teachers can contribute by adding comments with regards to the students' passages.	Formative
Worksheets	These materials are created to empower learning by revising previous knowledge that was instructed.	Formative
Essays	This is a method in which learners write about specific content. Teachers can ask students to write about any particular issue but it can also let students to decide about their interests in a certain area. Teachers can provide their comments about the language progress.	Formative/Summative
Self-/peer assessment	These two types of assessments can help students to track their learning trajectories. Learners can have the power to self-evaluate but also other students can assist to keep assessment more interactive and fruitful for the learners.	Formative

Teachers should be aware of the different applications and uses of other materials to enhance better learning practices inside the language classroom. The following section discusses the standard of communication of results in which teachers and other stakeholders should be involved to improve the assessment procedures in the classroom, as well as in their institution.

#### **2.4.3.3 Communication of Assessments and Results**

Communicating results in second language education is a notorious responsibility for language teachers. Teachers need to communicate students' results about their tests, progress and achievements. However, teachers should be aware that they communicate not only the scores of a test, but they need to anticipate learners about the content, method of delivering the instrument and any essential information related to the assessment before its final administration. Teachers should be objective when they are providing information about a testing situation, in other words, they need to provide how much percentage of the total grade certain tests or activities gain.

In the language classroom, the communication of assessment needs to be in three phases. The first phase is the anticipation of what students need to know about the upcoming test (e.g., contents, value of the total mark and method of delivery). Teachers need to be very explicit with regards to the value of certain tasks and activities, and at the same time how such instrument will be delivered. The second phase is once the students have taken certain tests and these results need to be given to the learners. It is important to have in mind that some students might feel anxious when tests results are communicated inside the language classroom. This is when teachers provide some input (feedback) to learners to identify what areas of their performances

need to be improved, but teachers need to highlight what areas students have mastered to make a positive progress. Feedback plays an important role at this stage. The last phase is to take the marks beyond the language classroom to communicate them to different stakeholders such as language programme managers, administrative officers or parents. Table 10 illustrates the three phases of communication of assessments with their applications.

**Table 10 Phases of Communication of Assessments (Brown, 2012)**

Phase	Purpose	Application
1. Informative	To provide the contents to be tested, the mark of the test and how this can affect students' final mark.	This is conducted before the administration of the test and it focuses to provide what the learners need to know about the forthcoming test.
2. Instructional	To identify what students have learnt and how students can improve their marks	This is after the administration of the test in which teachers provide feedback to students in the areas they have performed well and what they need to improve for the next testing situation
3. Impact	To communicate different stakeholders about the progress of the students	Once those scores have been communicated to students, the scores need to go to the academic office to visualise and track the students' progressions but also to let other parties know about these performances

In higher education, especially in Tamaulipas, the results are given to the students in two forms, face-to-face and via online in which parents and other stakeholders can have access to the institution website. Nevertheless, the instructional phase is the most important for the continuous performance of the students, teachers should be aware that the communication of results does not finish in the classroom. These results need to be observable and tracked either by other teachers from the same institution or the administrative officers. These could help understand and amend different perspectives when language programme managers can amend the learning, teaching and assessing methods in the curriculum.

#### **2.4.3.4 Educational Support through Assessment**

Once students have been communicated their results, teachers should be aware that these results not only affect or have a direct impact on the students. These results can assist teachers and language programme managers to modify and implement new methods for their teaching and assessing practices. Assessment can also bring a large number of benefits to language teachers and this is why it is important that teachers should be part of the process of developing language tests. Alderson and Wall (1993:120-121) lists 15 different hypothesis with regards to the use of tests and how these can affect teaching, these hypotheses are the following:

1. A test will influence teaching.
2. A test will influence learning.
3. A test will influence what teachers teach.
4. A test will influence how teachers teach.
5. A test will influence what learners learn.
6. A test will influence how learners learn.
7. A test will influence the rate and sequence of teaching.

## Chapter 2

8. A test will influence the rate and sequence of learning.
9. A test will influence the degree and depth of teaching.
10. A test will influence the degree and depth of learning.
11. A test will influence attitudes to the content, methods, etc. of teaching and learning.
12. Tests that have important consequences will have washback.
13. Tests that do not have important consequences will have no washback.
14. Tests will have washback effects on all learners and teachers.
15. Tests will have washback effects for some learners and some teachers but not for others.

These previous 15 hypotheses have been researched in different contexts and studies have highlighted that regardless the type of testing instruments, these tend to have some impact on both sides of the coin with teachers and learners. Bachman and Palmer (1996) and Popham (2011) stated that washback was an element that has major impact on summative assessment. Nevertheless, it is essential to understand the foundations of a testing instrument with regards to the position of this in the curricula. This is a practice that not many teachers do since several institutions do not allow the analysis of the tests once they have been administered. Another reason why teachers avoid exploring and examining students' tests is the lack of time when teachers must conduct this activity. It is essential to identify if the tests need to be amended or the methods of teaching practices. In Tamaulipas, teachers do not have certain opportunities to identify the impact of the tests due to the fact the test are returned to the administrative office. Table 11 summaries the assessment skills and component of school-based assessment.

**Table 11 Matrix of School-based Assessments Component**

Component	Standard	Skills and Practices
School-based Assessments	Language assessment	<ol style="list-style-type: none"> <li>a) Recognise the properties of assessment language receptive and productive skills.</li> <li>b) Use and adapt criteria for marking productive skills (reading and listening)</li> <li>c) Choose types of responses based on the purpose of assessments.</li> <li>d) Justify the use of assessment methods for linguistic features (e.g., grammar, vocabulary, etc.)</li> <li>e) Manage to assess linguistic and content properties based on the assessment methods</li> </ol>
	Assessment for learning	<ol style="list-style-type: none"> <li>a) Development of alternative assessments to promote learning</li> <li>b) Distinguish properties among alternatives (performance assessments) and standardised tests to prioritise learning</li> <li>c) Develop awareness of the advantages of portfolios and other assessment tools to collect learning evidence</li> <li>d) Prepare and predict students' performances based on alternatives for assessment, excluding language tests</li> <li>e) Align alternative assessments to a specific purpose of assessment (e.g., diagnostic, formative and summative)</li> </ol>
	Communication of assessment and results	<ol style="list-style-type: none"> <li>a) Inform students about the assessment practices and methods for marking before and during the course</li> <li>b) Communicate test-scores and other assessments marks to students</li> <li>c) Inform students about their progress (feedback) to promote learning</li> <li>d) Measure the impact of the tests towards students' learning trajectories</li> <li>e) Inform institutional authorities about students' lack of progress</li> </ol>

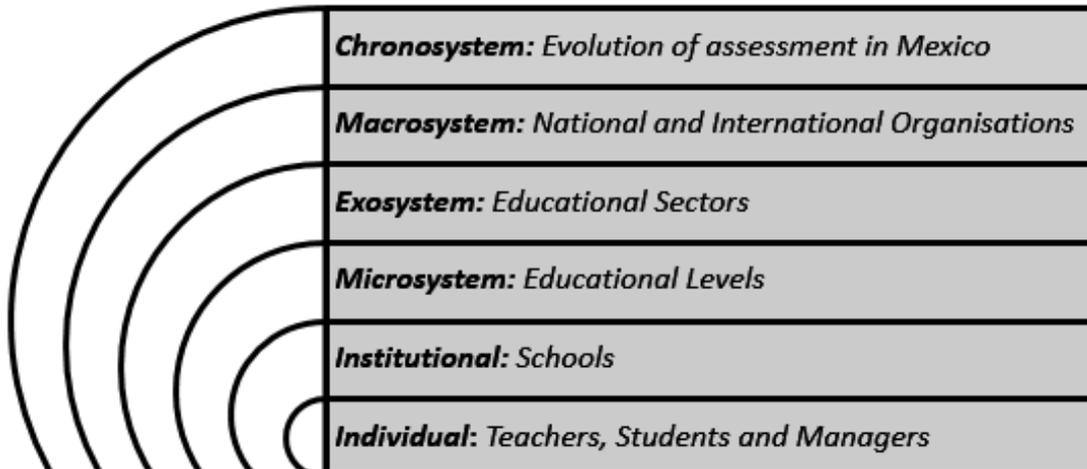
	Assessment as an educational support	<ul style="list-style-type: none"> <li>a) Compare test scores and other assessment to amend teaching, learning and assessing methods in the classroom</li> <li>b) Interpolate students' progress and teaching assessing methods to see efficiency in learning</li> <li>c) Examine and exchange students' performances to improve the quality of tests and other sources</li> <li>d) Recommend and suggest further development of assessment methods based on students' performances</li> <li>e) Evaluate the impact of tests for learning, teaching and assessing practices</li> </ul>
--	--------------------------------------	--

Exploring the knowledge of assessment, the technical considerations and the school-based assessments in terms of teachers' assessment skills and practices could allow the promotion of better opportunities for learning and teaching in the Mexican classrooms, especially in Tamaulipas. The following section discusses the review of the literature in terms of Mexican and Tamaulipas educational standards.

## 2.5 Contextualising Educational Assessment: Mexico and Tamaulipas

It has been suggested that it is relevant to examine the context in which assessments occur since language teachers, students and other stakeholders are mainly involved in this process. Assessment literates are individuals who are supposed to possess a repertoire of skills and practices to apply them adequately to improve the learning progresses when they are in the language classroom (Stiggins, 2002; Mertler, 2005). As a consequence, it is important place us in a time when English was introduced as a subject in a Mexican classroom. Historically speaking, English as a Foreign Language (EFL) programmes were firstly incorporated as part of the national curricula in several private secondary schools in 1925 in several states in Mexico, including the state of Tamaulipas. During this period, the Ministry of Education (SEP) had been active just several years, since it was founded in 1921 and Mexico was still weakened from the Mexican Revolution. From this point in time, the Mexican educational system became solidifying, but this took several decades to become fully recognise nationwide. Based on a comprehensive review of the literature about the Mexican assessment context, six aspects were considered necessary to as a possible influence to assessment literacy in Tamaulipas, Mexico. These aspects are 1) the identification on how educational assessment has been conducted in Mexico, so as to grasp different stages of evolution throughout time in the Mexican State of Tamaulipas (Chronosystem), 2) national and international organisations that promote educational support towards Mexican educational systems in general and towards the English as a foreign language field (Macrosystem) 3) the Mexican educational sectors (Exosystem) in which assessments and tests are administered (e.g., the public, the private and language institutes), 4) the educational level: institutions from the primary to post-graduate levels of education (Microsystem), 5) the institution, school and the classrooms, where assessments mostly occurs (Institutional), and finally 6) the individuals,

teachers and language programme managers, who have been developing and gaining certain knowledge, skills and practices during their professional trajectories. Figure 5 illustrates the six-layer diagram that highlights the analysis of how assessment policies are considered, not only in Tamaulipas, but also in the rest of the Mexican States.



**Figure 5 Multilayer Analysis of Mexican Assessment Policies in Education**

The most-outer circle represents the *chronosystem* that describes how individuals have dealt with their expectations on how assessment has changed throughout the years; however, this system also deals with the national decisions on the use of evaluation for selective, mobility and educational purposes. The evolution of assessment from the national perspective towards the individual might cause friction while making decisions of using assessments in different ways. The subsequent circle represents the *macrosystem* in which the federal government has reacted to certain educational reforms with the assistance of national and international support. Then, at the *exosystem*, those educational sectors that could have taken a strong position to determine how education works in Mexico. The *microsystem* includes the educational levels such as primary, secondary and tertiary education. This review discusses the upper-secondary (high-school system) and higher education studies. The *institutional* discusses the forms of assessments in schools and their policies. Finally, the inner circle, represents the *individual*, in this case, the teachers and how they take into consideration assessments in the daily basis. The following sections describe each of the prior components of the Figure 5.

### 2.5.1 Chronosystem

First and foremost, Dr. Martinez-Rizo has investigated the field of educational evaluation in Mexico and has contributed significantly to the national policies by providing great support to the methods of assessment in the classrooms and national standardise tests due to his publications and closeness to the Ministry of Education. In addition, he has highlighted that Mexican

educational system of learning is not currently at the finest stage, but improvements should be recommended (Martinez-Rizo, 2012). This researcher has outline three stages on how assessments in Mexico have evolved in over a period of five decades. Similarly, it is important to remember that De Luca, et al., (2010) also divided the period of assessment literacy evolution into three stages. The first period is from 1970 to the end of the 1980's. The second 1990- 2002 and the third from the 2002 until out current time (Martinez-Rizo and Blanco, 2010).

Between 1970 and 1989, SEP conducted a national scale project to improve the systematic collection of educational assessments by developing standardised test to cover the needs for selection and admission from primary to secondary education. Everything began in 1972, when it started the first administration of high-stake tests for measuring students' performances in basic educational levels with the intention to conduct diagnostic exams. Then two years later, the first institution of accreditation and evaluation was developed, and tests brought the first statistical reports on the performances from the selected students. However, this first attempt was not successful due to the lack of expertise from the policy makers in identifying crucial characteristics such as students' progression and achievements (Martinez-Rizo and Blanco, 2010). The official standardise test for educational purpose of admission for the basic education was introduced almost ten years later in 1982. Nonetheless, the institutional policies, the knowledge of assessment and evaluation, at that point in time, were almost inexistence. The lack of validity and reliability were two aspects that affect the transition of this towards educational purposes. It is worth noting that at the end of the 1980's, a notorious lack of knowledge, skills and practices in educational assessment from different stakeholders such as language teachers, policy-makers and other stakeholders from other regions from the world was noticed (Nunnan, 1989).

In Mexico, from the 1990 to 2002, significant decisions in evaluation and assessment were made by national and international organisations. For basic education, teachers became part of a sub-organisation developed by SEP, which rewarded teachers' professional achievement by administering standardised tests to their learners to measure how teachers had taught their students. 'Programa de Carrera Magisterial' or in its English translation "Program for Teacher Pathways", was created in 1994 to make primary school teachers aware of the importance of appropriate teaching and assessing methods in their education. In addition, the National Centre for Evaluation in Higher Education was also created (CENEVAL, acronym in Spanish) which was in charge for the development of the entry test for higher education studies. The first national measures for education were launched in 1998. This year, Mexico tried to progress to enhance and standardised educational assessments. In addition, at the end of the 1990's international measures were approved to be conducted in Mexico to compare and contrast their educational assessments. Other international measures were the TIMSS (Trends in Mathematics and Sciences

## Chapter 2

Study) in 1995, Latin-American Laboratory for the Quality of Educational Assessment (LLECE, acronym in Spanish) in 1997 and PISA (Programme for Institutional Student Assessment) in 2000 were incorporated into the Mexican Educational Systems.

At the beginning of the 2000's, the Mexican system of Education adopted a philosophy of using assessments to identify, prevent and improve the level of students' performances, the achievement reports on the administration and application from the institutional practices, to develop a professional development programme with a great deal of opportunities to empower teachers' practice and to plan a working agenda with improvement policies at the larger scale. In 2002 the National Institute of Educational Assessment (INEE) was officially funded by SEP with funds from the National Government. INEE developed several projects and reached to slightly increase the Mexican standards of education to higher levels. With the assistance of the Federal Government, SEP and INEE, Mexico was able to be part of the Organisation for Economic Co-operation and Development (OECD) among other 36 countries such as United Kingdom, USA and Spain, among others (OECD, 2019). INEE develop better strategies to operate institutional assessments, to promote communication of assessments in terms of results and access to materials to support the learning process and to use a diverse repertoire of assessment competences to empower educational support through assessment (INEE, 2012). In addition, a significant contribution from the SEP in this period was the design of a measures for their educational levels: basic and secondary education in 2006 with a further intention to be administered in upper-secondary levels. The Exámenes Nacionales de Logro Académic en Centros Escolares (ENALCE, the National Exams for Academic Achievement in Education Centres) aimed at collecting evidence on Mexican students' performances and these could be compared with international measures such as PISA or LLECE. Knowing these periods on how international and national examinations were emerging throughout time. This period, to my view, is when it became into the testing culture. The Chronosystem level is being discussed in the following levels as well. The following section discusses which organisations, nationals and internationals, have supported the educational systems in Mexico

### **2.5.2      Macrosystem**

Currently, there are four international organisations that have supported Mexico to improve their educational practices and promote new educational reforms that increase the national profile. As has been mentioned before, PISA scores indicated that Mexico is currently under the borderline scores required standards of education from the Organization for the Economic Co-operation and Development (OECD). The OECD is an intergovernmental organisation that promotes economic stimulate to empower national progress and world trade. This

organisation has made several recommendations for improving education by facilitating courses and economical resources, to prepare teachers during their pre-service and in-service phase. Also, to promote constant evaluation of teachers' performances and to increase professional development profiles. Another organisation that assist Mexican education is the United Nations (UN) that is the largest and most powerful intergovernmental organisation, which promotes an international co-operation among countries and it also deals with the international and economic relationships. These two international organisations guide educational policies to highlight the needs and how to improve education across different sectors and teaching contexts. Nevertheless, these new educational policies are ruled based on international standards that the government needs to meet, as well as monetary aspects.

Two main monetary sources allow governments to increase their status by providing loans that the government has the obligation to use effectively. The World Bank (WB) is the largest and most well-known development bank in the world that makes available loans to developing countries and the International Monetary Fund (IMF) which manages payment difficulties and international financial crises. IMF receives money from countries that are experiencing payment problems and they can borrow money. Mexican external debt has two-folded in the last 8 years reaching a historical debt of \$ 453, 548 million US dollars.

The previous four international organisations recommend, suggest and provide monetary sources and advice on improvements based on facts from national statistics (SEP, 2010). Moreover, these latter have also impacted on the field of education and language education. An association that was established for English teachers in Mexico entitled MEXTESOL (Mexican Association of English Language Teachers) was created in 1973 to promote the importance of the English language in Mexico and in its 32 States and guide teachers of English to have access to the most relevant issues in English language teaching. This association is a branch of TESOL (Teacher of English for Speakers of Other Languages) in The United States of America and IATEFL (International Association of English language Teachers) in the United Kingdom. Additionally, some states have their local branches to promote and discuss language issues into teaching practices and offer professional development opportunities such as academic sessions and regional conferences (MEXTESOL, 2009). Another important association that focuses on the performance of university teachers and students' progress is ANUPI (National Association of English Teachers at University Level). ANUPI was founded in 2003 with the intention to improve the quality of University teachers of English in Universities. This is a Mexican association that deals specifically with the competences of university teachers from all the country. MEXTESOL and ANUPI offer national conferences every year to present the latest resources, publications and to solve issues related to teaching English and linguistics. A variety of topics are presented to

## Chapter 2

facilitate the understanding of teaching. Many Mexican and International researchers attend these conferences to share and promote quality on the field.

Other International organisations have also supported several Mexican organisations to deliver international assistance to Mexican language teachers. The British Council has participated with many national institutions by preparing teachers from the National Programme of English in Basic Education (PNIEB, acronym in Spanish) and Universities. Also, there are publishing houses that provides continuing professional development to EFL teachers. MacMillan runs a special programme for English language teachers with different educational and interactive workshops, resources in different aspects of teaching, assessing and developing language resources. This programme is called RedNova and it is an online platform that offers courses every year for pre-service and in-service language teachers. Nevertheless, all these institutions promotes, not only assessment, but other areas of professional development. All these institutions are mainly in charge of the current educational trends of English Language Teaching in Mexico (Moore, 2017). It is worth noting that the courses, training and workshops that are provided by these institutions are sometimes linked to areas such as knowledge about the subject, managing students and materials for learning and teaching; assessment, on the other hand, the number of assessment courses and workshops in conferences in Mexico is currently limited.

### **2.5.3 Exosystem**

All Mexican citizens have the right to study and develop academically. The Mexican educational system is divided into two main sectors: the public and the private. All educational institutions are regulated, fully or partially funded, and certified by the Mexican Ministry of Education (SEP). Each sector has features that differentiate from each other. These features cover a large number of aspects such as time of instruction, the socio-economic status of the institution, curriculum and development of materials for learning, teaching and assessment interactions, etc.

Public institutions are the ones that have national agendas for the teaching and learning process and these mainly rely on the Ministry of Education. This sector mainly structures its time of instruction and learning in two class timetables: one in the morning and the other for the evening classroom. On the other hand, private institutions work autonomously, and they can develop their own timetable. This previous sector usually expands their timetables since they include extracurricular activities such as sports, modern languages and cultural activities. The public institutions tend to have a large number of students because this sector has, to some extent, the obligation to respond to the social demands to allow learners to study in their institutions, mainly for basic and lower secondary education (SEP, 2011). It is important to highlight that these institutions are to an extent free of tuition fees, but administrative fees need

to be paid annually by the students. This is the opposite for the private sector, the majority of this sector's classrooms have a limited number of students since it is believed that the cognitive process might occur faster due to student-teachers interactions (Roux & Anzurez Gonzalez, 2015). In relation to the socio-economic status, the private institutions charge a tuition fee and administrative changes. Students are required to pay these fees monthly.

Although teachers in both types of educational sectors, the public and the private, should teach and deliver the same content; students' cognitive process can be different based on the curriculum delivery and access for cognitive development. The quality of teaching and students' demands on the development of materials are possible to differ from these two educational sectors (Martinez-Rizo, 2012). The public institutions offer a curriculum followed by the National standards proposed by the SEP and the INEE. The private sector has the flexibility to modify the learning-teaching methods, but the contents would reflect the same conditions from the public sector. Not a large number of the public institutions offer several extracurricular activities or workshops that are complementary, but private institutions have the obligation to provide high quality learning experiences. These institutions also have a significant infrastructure to develop knowledge and expand the curriculum so learners could achieve a number of credits related to artistic, sport or musical performances (SEP, 2011). Institutions that belong to the public sector provide their own course books free of charge across all basic and secondary education years. The learning materials and course books are developed by the SEP. Learning materials for the private sector are not included as part of the administrative fees, therefore, students are required to obtain these materials for their modules (SEP, 2008). The quality of teaching and learning rely on how the institution and teachers work collaboratively to promote better learning resources and to increase the institution profile. According to the State Government Law for Education, the public sector should deliver teaching and recruit students regardless of their religious backgrounds, socio-economic status and must be mandatory for every Mexican. A large number of Mexican private institutes follow a catholic-Christian doctrine, which religion is part of their curriculum and has a significant impact on students' behaviour and performances.

In terms of language education, private institutions have been well-known for their institutional interest in teaching modern languages such as English, French, etc. these institutions offer sociocultural and linguistic courses in which learners can develop cognition based on the integration of social skills and culture. English is formally taught from pre-school to university and it follows international standards to constantly assess students' linguistic ability. The public institutions seen modern languages, mainly the English language, as part of their curriculum and is a national requirement since the Basic Education dissemination in 1993. Based on standardised

tests, the private institutions tend to achieve higher scores based on measurement National Tests such as EGEL or CENEVAL (OECD, 2012).

Emerging from this review of the literature, there is another sector that should be taken into consideration to understand EFL teachers' language assessment literacy in Tamaulipas. Language institutes are sometimes language schools or centre for languages that are directly linked to a public, private or independent institution that offers the services of providing modern language instruction. This type of institution regulates, adapts and implement strategies to promote the curriculum for foreign languages. As an illustration of a language institute in the State of Tamaulipas, the Centre for Languages and Applied Linguistics <sup>8</sup>(CELLAP) is a language school that is incorporated to the Academic Secretariat from the Autonomous University of Tamaulipas (UAT, 2006). This public institution, UAT, has 3 CELLAPs around the State of Tamaulipas providing language courses to university students and external students. CELLAP is also a centre that promote professional development throughout teaching qualifications, mainly from Cambridge English Assessment Examinations and the Educational Testing System. Other example of language institutes is the Iberoamerican Institute for Languages that in Spanish is known as 'the Institute Iberoamericano de Idiomas' <sup>9</sup>. This is an institution that has two functions: on one hand to promote general education from primary to upper-secondary studies being a private institution and on the other hand, to support language education practices. Therefore, we can assume that there are three educational sectors that promote language learning in different institutions. Assumedly, these three sectors might operate assessment differently; as a consequence, it is essential to examine what are the similarities and differences among these sectors regarding their teachers' language assessment skills and practices. The following section provides the microsystem which deals with the educational levels.

### 2.5.4 Microsystem

Mexican education is divided into three levels. The first one is Basic Education and this one is divided into three sublevels i) Pre-school (ages 3-6), ii) Primary Education (grades 1 to 6), and iii) Lower Secondary Education (grades 7 to 9). The second level is Middle High Education and this one is divided in two sublevels, i) Technical Professional and ii) Baccaalaureate. The last level is

---

<sup>8</sup> The CELLAP (Centro de Lenguas y Lingüística Aplicada) is a centre that belongs to the Autonomous University of Tamaulipas (UAT). This language institute offers learning instruction from a range of languages such as English, French, German and Italian courses.

<http://www.cellap.uat.edu.mx/SitePages/principal.aspx>

<sup>9</sup> The Instituto Iberoamericano de Idiomas is a language school that also is part from Basic and secondary educational system. This is a private institution, but it has its own centre for languages.

<https://institutoibero.com/>

higher education, and this is divided into three sublevels, i) High Technician, ii) Bachelor or Undergraduate, iii) Postgraduate, Master Degree or Doctorate Degree.

In order to be admitted in the Middle High Education (upper-secondary schools) all Mexican students are required to take the IDANIS (Instrument for the Diagnostic of New Lower Secondary School Students) and EXANNI I (National Upper Secondary Education Entrance Examination). IDANIS and EXANI I, two national measures, test knowledge from different areas that were taught in Basic and Lower-Secondary and these are administered by SEP. The Education curriculum is also designed by SEP and contains a great deal of modules to guide students to become independent to choose one discipline for their higher education studies. In this educational level students are becoming autonomous and they are responsible for their own educational choices. The number of subjects and credits are significantly larger than basic education (primary and secondary) due to the importance of the specific subjects that students desire to lead their fields. Some of the modules that make difference are Applied Mathematics, Chemistry, Physics, Biology, Ethics, Technology, Electronics, Accountancy, Modern Languages, etc. Students are awarded with two different certificates i) Certificate of Lower High Education and ii) Certificate of Middle High Education or Certificate of Technical Practitioners.

On the contrary, higher education institutions do not have a national syllabus, as the Basic and Middle Education levels, since there is a large diversity of institutions that focuses on different disciplines; however, SEP awards all successful students who achieve satisfactory grades and standards from the institution. SEP also created two Examinations for the Admission and Graduation purposes. In 2012 there were more than 2,500 Higher Education Institutions in Mexico and all of these belong to different subsystems of universities. This research focuses on two types of Higher Education institutions: Public and Private University subsystem (SET, 2018).

On one hand, approximately 65 federal and state universities in Mexico form the Public University Subsystem. Some of these universities are considered as autonomous institutions which they possess a wide extent of power in the development of the curriculum, organisation and evaluation of the programmes. On the other hand, the Private University Subsystem lists more than 2,000 institutions; nevertheless, the validation and regulations may remain on one autonomous university due to the importance and corporation with the SEP. All type of universities and institutions can provide a large number of disciplines and degrees only in they have been approved by the SEP regulations due to the fact this issues the Certificates of Higher Education.

The evaluation criteria range, or measure as previously discussed, in mostly all Mexican institutions is from 0 to 10: Higher educational institutions, on the contrary, tend to use a

## Chapter 2

criterion range from 0 to 100. The minimum mark to satisfactorily pass is 6 or 60 while the maximum is 10 or 100, and from 0 to 5 or 50 stands for fail or not accredited. These ranges are based on how institutions decide their own marking policies. The methods of assessment for all Mexican institutions remain on formative and summative. Evaluation is an important issue since the creation of the INEE (Institute for National Evaluation in Education) due to the constant assessment of programmes, institutions and teachers in order to nationally standardise the teaching, learning and assessing process in the country (INNE, 2005).

Then, in 2008, English was included in high schools and technological bachelors as a mandatory programme. In 2009 the (Integral Reforms for the Basic Education) RIEB and the National English Programme for Basic Education) NEPBE were incorporated and they dealt with the better professionalism of the teachers in institutions where English is taught (SEP-SNTE, 2008). In 2011, the National English Programme for Basic Education was established for all basic education including primary and secondary schools (SEP, 2010). It was not until 2012 that English was incorporated at the tertiary level in polytechnic universities. A new reform named as the Bilingual, International and Sustainable Universities (BIS) was recently developed in 2018 and the SEP developed the National English Strategy for Normal Schools in which deals to prepare English language teachers across the country. Mexico has incorporated several reforms for the National Strategy of empowering the English language across the country. However, these reforms cannot be taken into consideration if teachers were not eager to empower their language teaching practices into their institutions. The following system is the institutional level where teachers need to attend their schools or organisations to deliver the instruction.

### **2.5.5 Institutional (The School)**

In order to understand how tests and assessments are conducted at institutional level, it is worth noting the role of the teachers and institutional authorities towards educational assessments is vital. The assessment processes ideally occur in a classroom, but this condition might be influenced by the institutional assessment policies concerning to standards of the institution (Inbar-Lourie, 2008b). Teachers and programme managers usually operate this educational assessment practices with the intention to track students' progress, identify what students have learnt and deliver feedback to students to improve further performances (McNamara, 2008). The role of teachers' assessment skills and practices might be different depending the situation where these occur.

Assessment practices are inclined to be stipulated by the educational authorities and how teachers conduct such assessment practices. It is my belief that teachers sometimes have restrictions at the moment of administering tests and classroom assessments. Assessments ought

to be seen as tools to enhance the relationship between teachers and students to achieve learning and teaching processes. However, institutional authorities work collaboratively or independently with teachers to enhance the process of teaching, and to maximise the assessment educational methods to achieve the purpose of learning in the classroom. Literature has highlighted that not all language teachers develop their own tests, and these could be given by their institutional authorities or a lack of language tests can be the case (Sikka, Nath & Cohen, 2007; Popham 2011). I have included the following three scenarios, which were developed based on my own teaching experiences, in which teachers could deal with language test and other assessments.

**Scenario 1** *“I am responsible for four EFL classrooms with 39 students in each. I develop my own language tests and develop some classroom activities that are part of the evaluation. The evaluation consists of 60% of the tests and 40% in other types of assessments as portfolios, worksheets, projects, etc. I don’t administer speaking tests because of the lack of time. My students tend to get very low marks in their tests but they pass because of the alternative assessments. I deliver feedback in small groups but individual feedback is impossible”*

**Scenario 2** *“I have 3 EFL classrooms and there are approximately 16 students in each. I do not develop the language tests; these are given to me by the language programme manager. I do mark the tests and I have to provide feedback. The students have to meet the minimum mark which is 7/10 to successfully upgrade to the next language course. Students have some coursework but their performances are not reflected on the final grade. Students tend to get low tests-scores and most of them, unfortunately, do not pass the language course”*

**Scenario 3** *“I work full time and I have 5 EFL classrooms with approximately 22 students in each. The tests are given to me by the coordinator of the programme. The institutional assessment criteria consists of 50% summative tests and 50% formative. I use portfolios with my students so they can collect learning evidences to measure their progress. Students also use course-books and I frequently select certain tasks from the workbook so students can complete and be assessed. However, my students do not always complete several tasks, but they meet the minimum mark to pass”*

The assessment practices of these three aforementioned scenarios differed among them; not only on their educational sectors, but the way that each institution operates their assessment practices. It is seen that other stakeholders (i.e., local authorities, language programme managers, coordinators, etc...) were involved in making decisions on their students’ assessment processes. The rationale behind of integrating other stakeholders to the process of assessment is to validate and to contribute, either to the improvement of students’ learning through the use of standardised tests, or to increase the institutional profile.

View on a superficial level, scenario 1 was the only teacher who developed the classroom tests and specified their own criteria. This teacher could be a good example of teachers’ *full institutional operational assessment systems* since this teacher oversaw the developing of tests and other sources to assess students’ language proficiency. Teacher from scenario 1 could make

## Chapter 2

judgmental decisions to upgrade the students to the next language learning stage without informing or reporting to external authorities. The other two, scenario 2 and 3, did not develop their own classroom tests as these were provided by their institutional authorities. In a testing perspective, these two teachers have, to a certain degree, a *limited operational access* to assessment. Scenario 2 only used the testing resources that were provided by the institutional authorities and no other assessments practices in the classroom were recorded. Teacher from scenario 3 did not develop the tests and these were similarly provided as scenario 2; however, scenario 3 used other resources to assess and to track students' progress and was responsible to provide feedback to students. For scenario 3, the institutional operational assessment system becomes partial.

These three scenarios highlighted that either teachers and/ or programme managers tend to be responsible for the assessment processes and practices of the students. These three categories of operational assessment in institutions can assist to discover what teachers really required to develop their assessment skills and practices. I am now presenting a description of these three institutional operational assessment systems and what the roles of teachers and programme managers are within their institutions:

- *Full institutional operational assessment system* refers to the situation in which language teachers are in total control of the test development and administration of other assessment sources to be administered to their students. The institutional authorities are not involved in the process of assessing students. Therefore, teachers can make their own final decisions of upgrading their students based on their progress and test scores.
- *Partial institutional operational assessment system* refers to the situation in which language teachers and LPMs are in control of the test development and/or administration of other assessment sources to be administered to their students. The institutional authorities are involved in the process of assessing students by developing testing sources to standardised, to some extent, the teaching and learning processes. Therefore, teachers make their final decisions of upgrading their students based on the tests developed by the institution and the assessments conducted on the behalf of the teacher.
- *Limited institutional operational assessment system* refers to the situation in which language teachers have no control of the test development and/or administration of other assessment sources to be administered to their students. The institutional authorities regulate the complete process of assessing students' proficiency by developing or adapting testing materials. Therefore, teachers do not make their own final decisions of upgrading their students. It is under the institution solely that decides finalises students' test-scores and assessment policies.

## 2.5.6 Individual (The Teacher)

Research studies have identified that teachers' experiences, academic and professional trajectories have influence teachers' assessment skills and practices. In this section, these factors are discussed and presented with a comprehensive review of literature and empirical studies to highlight a further need of studies. This section begins with the term experience to understand the importance of this feature towards the appropriate assessment purposes and practices. Then, the practitioners' academic and professional trajectories are discussed along with the concept of professional development and its opportunities.

### 2.5.6.1 Previous Assessment Experiences

The term 'experience' encompasses a broad range of aspects. Experience has been defined as the '*process of getting knowledge or skill from doing, seeing or feeling something*' (Cambridge Dictionary, 2015). Since this is the process which describes the action of any experience, we can assume that experiencing something builds up an acquisition of knowledge and skills by perceiving and evaluating its unfamiliarity until its being gained. However, in education, a learning experience refers to those situations in which practitioners identify good and bad properties of something being learnt (Richards and Nunan, 1990). Experiences play a crucial factor when we are learning something new. I believe that this connotation of experience could also be placed at the moment we are in the classroom as language teachers, where we are facing constant challenges to improve the quality of instruction by implementing new skills and practices. Teachers are the one who have been in two situations, when they were learners and when they are instructing the language in a classroom. Teachers require to have the ability to identify the quality of an experience, either positive or negative through self-reflection. Teachers are currently developing knowledge and techniques in the classrooms that allow them to connect schemata to model new practices through confidence since they become aware and, to some extent, critical when something in their teaching practice works well or not (Nunan, 1992).

In relation to teachers' assessment skills and practices, previous experiences about classroom assessment play an important role. With the intention to explore teachers' previous assessment experiences, it is considered important to scrutinise teachers' previous experiences using assessments in the language classroom as language learners and teachers' current experiences of using assessments. Previous assessment experiences as language learners could lead to a better teachers' assessment skills and practices in the classroom. Language teachers are responsible for conducting different practices in the classroom, with most of these practices being applied because they believe, to some extent, these are appropriate. Research has shown that teachers' assessment skills and practices mostly relied on their previous experiences as language learners, which have influenced the current selection of assessment methods and instruments

## Chapter 2

(Quiter and Gallini, 2010; Mussawy, 2009; Sikka, Nath and Cohen, 2007, Barns, Fives and Dacey, 2017).

Quiter and Gallini (2010) investigated teachers' attitudes towards assessment literacy. In their research study, they presented a comprehensive literature review and highlighted a diagram that demonstrated the relationship between assessment literacy, personal experiences and current attitudes system. The results from this study demonstrated that teachers develop knowledge based on their experiences and attitudes, since these could enhance new practices in the language classroom. They also stated that teachers commenced building their professional practices from the moment their language learning journey started. This suggests that whilst prospective teachers started learning the language, especially for those whose English is not the first language, this could identify which methods were the most suitable for studying and revising the language.

By the same token, throughout our learning journey we have had different tests and assessments that we could consider which ones were more suitable, effective and the impact they had on our learning trajectory that might influence the way we conduct assessments. Similar findings were found by Mussawy (2009) who developed a questionnaire and asked participants (n=207) about their assessment experiences between teachers and students. The results showed that there was a positive attitude towards the assessment practices because of how learners could provide teachers with the application of new techniques in the classroom. The quantitative researched conducted, the survey findings, presented a methodological limitation stated by the author who highlighted that the participants' first language, in this case not English, may have influenced the outcomes of the study. Slightly different outcome was the study by Jannati (2015), who conducted a qualitative research through semi-structured interviews with Iranian teachers (n=18). The purpose of this research was to explore the teaching' experiences dividing participants in three groups based on their language assessment literacy levels. Findings showed that participants' experiences did not play an important role on how teachers administer current assessment skills and practices, but their experiences could allow them to identify those methods for certain groups of students. A limitation of this study was that all participants were previously registered in a MA TESOL programme since this could have influence how participants perceive their assessment experiences.

A relevant contribution about how teachers perceive assessment based on their experiences was a study conducted by Coombe, *et al.* (2012) who identified four features that affect teachers' experiences and language assessment literacy. The authors conducted a study to explore the field of assessment literacy in relation to teachers' experiences of conducting

assessments in the language classroom: these are 1) a lack of knowledge and fear towards the subject, 2) specialised knowledge of testing and assessment concepts, 3) a lack of test development experiences, and 4) a lack of time for assessment skills and practices.

Language teachers require to have adequate linguistic and theoretical level of the target language since they are teaching it, in this case English, to accurately identify learners' language gaps throughout their performances, coursework and in learning resources. Borg (2006) indicates the importance of teachers' learning experiences in their second language could be crucial to their teaching practice. These experiences could contribute to a negative perspective if the learning process was not satisfactory. In other words, language teachers should have a lexical repertoire appropriate for the courses they are in command.

The second is the lack of knowledge teachers required to assess language learners successfully. Testing and teaching are two essential aspects for the learning process. Sheehan and Munro (2017) pointed out that the testing resources for teachers are not always accessible due to financial restrictions and technicalities. They complained that language teachers should have access to the research that has been conducted lately to motivate others to identify their language assessment literacy and to enhance their practices through reflection and empower their skills in the language classroom. The lack of testing and assessment modules in educational programmes that prepare language teachers could influence the way in which assessment is perceived by these pre-service teachers. Ward and Murry-Ward (1999) pointed out that teacher training programs include a large number of elements that are essential for a basic educator, but sometimes these courses do not cover what teachers really need for their performances and inclined to fail in developing adequate testing and assessment sources.

The third problem is the lack of opportunities for teachers to develop purposeful tests as the testing industry has gained power in developing, on a large scale, materials from different publishers (e.g., Cambridge University Press, Macmillan, etc.) have designed standardised tests and mock tests that are usually aligned to the contents of the course books. Teachers believe that is beneficial to use these testing resources since they have been validated and they are suitably created. Impara, Plake and Fager (1993) and Marso and Pigge (1993) mentioned that these testing materials from course books might not always have a direct impact on how students learn since each learner is individual and each one has their own learning needs.

The proper development of testing instruments has an important effect that is named as washback, previously discussed above as a testing principle. By analysing the impact of the test, it is possible to identify whether the materials were suitable for the learners. Besides, it can also help to guide teachers to amend the methods of teaching to improve learning. Coniam (2009)

## Chapter 2

indicates that tests created by classroom teachers are sometimes poorly designed in terms of quality and these conditions do benefit neither the teacher nor the students. Tests need to be revised to meet the requirements for the testing development process and the institutional intentions. Teachers could improve their materials if these ones are revised before being administrated. Moreover, in this in-depth revision, teachers can obtain two important factors: 1) identification of how students' best learn a language, and 2) link teaching and assessing to improve the quality of the instruction and learning.

Final, the fourth problem is the lack of time that teachers have to spend assessing their language learners. Teachers have a significant number of responsibilities in the language classroom; however, assessment is one aspect that takes most of the teachers' time in the language classroom (O'Sullivan, 2017). Assessments needs to be properly planed, designed and revised in advance in order to provide feedback to the leaners accordingly to their needs and the contents of the course. Unfortunately, feedback can be a process that takes a lot of time.

Teachers' experiences play an important role especially how they choose and adapt other instruments towards the purpose of classroom assessments. This section included teachers' experiences as learners and teachers. However, one research study highlighted that the academic discipline might have a positive impact towards the development of language assessment literacy (Jannati, 2015). The following section includes how practitioners' trajectories, language teachers, could influence teachers' language assessment literacy by discussing academic, professional and professional development.

### **2.5.6.2 Teachers' Professional and Academic Trajectories**

Teachers' professional and academic trajectories are vital element to explore due to the importance that brings to understand how teachers use a repertoire of practices and skills in the language classrooms. A large number of language teaching programmes cover several topics and contents related to the field of English language teaching. Educational programmes for pre-service teachers are sometimes equipped with different areas that focus on how to teach language lessons, knowing the learners, etc. Unfortunately, the content for language testing and classroom-based language assessment in some of these educational programmes is superficial (Nunan, 1992; Fulcher, 2012).

Since 1995, the English language teaching field in Mexico has been forced to develop new programmes to prepare pre-service and in-service teachers across the country due to the importance of the English language and the recommendations from the OECD. The British Council in Mexico and SEP introduced the English Language Programme in Basic Education during the 2000s. A lot of work has been done to improve the quality of language education and teaching

programmes. This programme has been incorporated into primary and secondary schools. The English language is now mandatory, and learners need to achieve certain linguistic level during their studies. This national movement for the development of language education in Mexico has been the centre of attention to preparing English language teachers with better opportunities for their teaching practice.

The number of students taking English lessons across the educational sectors has forced authorities to explore and to provide better opportunities for their learners. Mexican students need to take several English language examinations for several reasons. The British Council Mexico conducted a study to identify the needs of learners with regards to the English language. Data showed that 50% of the participants pointed that they learn this language for seeking better job opportunities. Data also showed that higher education students tend to take TOEFL and IELTS as part of their language Courses. Mexico is a country that has been guided by international language certifications. The English language testing market has rapidly increased in the last 20 years in Mexico, and this has been a beneficial aspect for universities to increase their profile for internationalisation.

The Ministry of Education, SEP, has approved to recognise two world-wide teaching qualifications from the department of Cambridge Assessment: the Teaching Knowledge Test (TKT) and the In-service Certificate of English Language Teaching (ICELT) as the minimum requirements for professional practice in all educational levels in Mexico (SEP, 2010). However, research has shown that apart from taking and successfully passing these teaching qualifications, the reality seems to be different once the teacher is in the language classroom. Recently, the British Council started online master and doctorate degrees in English language teaching with the University of Southampton to increase and benefits for those English language teachers who desire to improve their professional development and they cannot study a face-to-face degree. A large number of Mexican practitioners have been awarded these degrees to expand the national language education profile to a higher level.

Additionally, the British Council Mexico in collaboration with the University of Southampton published a book with articles from Mexican PhD students (Grounds & Moore, 2017). Moore (2017) conducted a study to analyse the current situation of Mexican teachers of English in higher education. She observed the evolution of the English language teaching field from the last 25 years in this country. Her research focuses on the Mexican teachers' perspectives towards the use and application of English language teaching and learning, but one component of this investigation is, assessment in the language classroom. She discovered that assessment and testing practices are not appropriate due to the lack of interest from the teachers. Mexican

## Chapter 2

teachers tend to be in front of a classroom a large number of hours. Teachers have different responsibilities, and this study explored the reasons why teachers are not interested in improving their methods to assess learners adequately. Results showed that teachers spend many hours planning their lessons, teaching and marking students' tests and coursework. The quality of the assessment and testing practices has been identified as poor for the lack of time teachers have to create instruments.

Moore also made emphases that the public sector is the most affected in the Mexican context. Numerous students in the language classroom have been affected the performance of teachers' professional development. Besides this, teachers are not only affected but also students. The quality of marking piles of essays and tests might be significantly weak compared to those classrooms with a small number of students. The private sector, on the other hand, offers a variety of courses with a lower number of students. Private higher education institutions are well known for seeking more top ranks of educational systems. This is a major factor that inspires Mexican learners and teachers to have better opportunities to improve their careers. In addition, Gonzalez (2017) conducted a study, in the state of Tamaulipas, to explore the perspectives and level of training about language testing and assessment in the area of assessing writing. She found out that teachers from private and public higher education institutions were not well prepared with the essential knowledge they need to complete this task. Training courses to improve the assessment practices are essential to facilitate better teaching and assessing methods.

It has been suggested by a large number of studies that providing mentoring or teacher development courses in the institutions can provide plenty of advantages for teaching performances. As a result, teachers may not be aware of the specific foundations that institutions aim to achieve. If these professional development courses are provided, teachers can be aware of the teaching and assessing methods and needs that are required to achieve students' necessities. A major issue that can be seen in some institutions is the fact that some educators do not have the essential background for the courses they are responsible for. In Mexico, this situation is constantly occurring since institutions have a need to hire professionals with a multidisciplinary background. In several institutions where English is taught, a language educator or a professional with another degree, which is not related to the core of the subject, can teach this module.

Another important fact that has been presented in the Mexican teaching context is the lack of motivation and time to develop new materials for testing that can capture students' progress and tend to rely more on the educational context. There is a vast number of institutions that are using electronic materials that are mainly provided by the course books that teachers used to promote learning in the classroom. Most of these English language-learning materials offer a

variety of resources to assess students' knowledge and progression. Teachers use these materials without having in mind if those resources are assessing what the learners need in order to successfully be promoted to higher levels.

Professional development is defined as an extension of the academic trajectories to empower current practices with the assistance of reflecting, working in collaboration with other teachers and institutions (Richards and Farrell, 2005). The British Council developed a framework for teachers' professional development. This framework is named as Continuing Professional Development (CPD). Table 12 shows, on the left side, the 12 components of this framework and the four stages of development, on the right side. The 4 stages of development work with each of the components. In addition, this framework also includes how teachers should progress throughout the professional development with regards to their language proficiency, educational qualifications and English language qualifications.

The component for assessing learners is build up with 10 elements: 1) applying principles to measure learners progress, 2) using a range of assessment tools, 3)use assessment to inform teaching practice, 4) defining assessment criteria, 5) analysing learners' mistake to deliver constructive feedback, 6)use self- and peer assessment to develop their assessment skills, 7)preparing learners to take standardised tests, 8)sharing assessment results to different stakeholders, and 9) administrating, marking and collecting assessment data, and 9) reflect on the effectiveness of the learners progress.

**Table 12 Continuing Professional Development Framework (British Council, 2015)**

CPD Framework Components	Stages of Development
1. Planning lessons and courses	1. Awareness
2. Understanding learners	
3. Managing lessons	
4. Knowledge about the subject	2. Understanding
5. Managing resources	
6. Assessing learners	
7. Integrating technology	3. Engagement
8. Responsibility of personal professional development	
9. Using inclusive practices	
10. Using multilingual approaches	4. Integration
11. Promoting 21 <sup>st</sup> century skills	
12. Understanding educational policies and practices	

The element of assessing learners comprises of several language assessment literacy standards, but teachers need to understand different aspects that were not included. For language testing and classroom-based language assessment, it is essential to understand the development of testing materials, analysing the efficiency of those instruments with basic statistics that can report on the quality of the instruments. Other important consideration is the

fairness and ethics of assessment; these focus on how to make instruments fair to all the students' needs and how teachers should identify and keep the assessment data (results) confidential following ethical procedures. This framework is a relevant example that other fundamental elements of assessment need to be taken into consideration when teachers increase their professional development. It is important to mention that the knowledge the teachers acquire throughout practicing in the classroom, attending conferences and workshops, achieving higher educational degrees and teaching language certifications can lead to an outstanding performance as a teacher.

Teachers are required to continue their professional development throughout different sources such as testing and assessment books, teacher training courses, online resources such as videos, webinars and conferences. Nevertheless, there is a lack of materials for development of language testing and classroom-based language assessment. Taylor (2009) identified that the materials and resources that educators are using to teach how to assess language, are not appropriate with the essential features that teachers require to know with the intention to assess learners adequately. Malone (2013) conducted a study to create online materials for language teachers and test developers to improve the quality of language tests and testing instruments. The materials were validated and reviewed by different institutions. All these materials were chosen based on data that was collected throughout a survey in the United States. Likewise, Lukin et al. (2004) examined the contents of several teacher preparation courses and they pointed out that language testing and assessment is not as important as other components of professional development.

This section (see Section 2.5) provided an ample contextualisation of the Mexican education and educational assessment. A total of six levels of analysis were presented to explore the evolution of educational assessment in Mexico and Tamaulipas, the national and international organisations that support educational practices and the English language, the educational sectors like the public, private and language institutes. Also, the educational levels from basic education to tertiary education, the institution where different agents work collaboratively and finally, the individual level where the teacher plays a fundamental role in educational assessment. The following section provides a review of the literature of empirical studies on language assessment literacy.

### **2.6 Review of Empirical Studies on Language Assessment Literacy**

Research in the field of assessment literacy has been developed since the American Federation of Teachers identified a notorious lack of teachers' understanding and comprehension of testing and assessment in education (Stiggings, 1991). As a result, Plake and Impara (1993)

achieved to firstly design a questionnaire based on the framework of assessment Literacy which was proposed by the AFT in 1990. This questionnaire was labelled as “Teacher Assessment Literacy Questionnaire” (TALQ). The purpose of this instrument was to gather information about how knowledgeable American teachers were. TALQ was administered in 98 districts in 45 American States to identify teachers’ assessment literacy, but at the same time, to tackle current situations that teachers need to adapt to improve their lessons. Results report on that the level of assessment concepts knowledge was slightly low. Years later, Campbell and Mertler (2005) renamed the latter instrument and slightly modified and renamed this questionnaire as “Assessment Literacy Inventory” (ALI). This instrument was administered to 394 pre-service teachers in two terms between 2003 and 2004. The essential purpose was to validate the instrument based on Stiggins’ contributions, years after Plake and Impara developed the first attempt to evaluate teachers’ assessment literacy. ALI was analysed throughout item analysis and reported on an overall reliability index (KR20= 0.74), which experts in research indicate that an index above from 0.70 is a reliable instrument (Cohen, Manion & Morrison, 2011). Differences between these two instruments were highly on the feature of assessment for learning. These instrument TALQ and ALI have a time gap more than a decade. Alkharusi (2011) conducted a study to analyse the psychometric properties with Omani teachers 251 pre-service teachers who were undertaking three different undergraduate programmes. The methods employed by this study were item analysis, difficult level discrimination level and validity and reliability. Results report on that TALQ was an appropriate tool for the study and research on the educational assessment literacy. Some of the main changes from this period was the inclusion of several factors that led to an improvement of assessment skills and practices such as assessment for learning, ethical considerations and the use of psychometric properties. However, during this period, another instrument was developed with an extensive repertoire of assessment competences.

The instrument developed by Zhang and Burry-Stock (2003) analysed and explored the practices and skills throughout an instrument labelled ‘*Assessment Practice Inventory*’ (API). Participants (n=297) from different educational levels and subjects took part of this research. Results showed that teachers with higher level of training are more likely to use a large number of assessment practices. The researchers also suggested that frequent training is essential to motivate teachers to use an extensive repertoire of assessment practices. The latter instrument API had a dual purpose of data collection; the self-perceived knowledge of teachers and the frequency they use assessment practices.

Similarly, the *Assessment Literacy for High School Foreign Language Teachers Inventory*, which is an instrument developed by Newfields (2006) with 70 items related to assessment issues.

## Chapter 2

This instrument was divided into four parts: terminology in educational assessment, procedures, test interpretation and ethical aspects. This questionnaire was administered to three groups of stakeholders in Japan: 1) university undergraduates in educational programmes, 2) high school foreign teachers, and 3) test developers. Results showed that this instrument has a large number of high-quality psychometric properties and it assesses a specific educational context, but provided a superficial perceptive on how the teachers were capable to understand assessment and its purpose towards university undergraduates.

Quilter and Gallini (2000) conducted a study to explore the teachers' assessment literacy and attitudes. The participants (n=117) from this study were from the school district of Michigan. Two instruments were essential for the purpose of this study. The researchers used the TALQ by Plake (1993) to identify participants' assessment literacy levels. Apart from this, an instrument developed by Green and Stager (1989) and Green (1992) was administered to teachers to explore the attitudes towards language assessment. Results showed that personal experiences play an important role when language testing and assessment occur in the language classroom. This study contradicts the majority of the studies related to language assessment literacy since the authors pointed out that professional development might, to some extent, contribute to a better understanding of assessment in the classroom, but it is not the only vital element for teachers.

In relation to professional development and language assessment literacy, Sikka, Nath and Cohen (2007) conducted case studies to identify the importance of language assessment literacy and the professionalism of language instructors (n=4). Their findings show that assessment is a crucial factor and it is low supported by the institutions. It also showed that tests are frequently administrated with the intention to meet the national standards, but they do not focus on the process of learning. In addition, participants from this research pointed out that the high-stake tests do not promote effective language learning and they lacked knowledge and use for assessment tools in participants' teaching practice.

Wang, Wang and Huang (2008) conducted a study to identify the assessment practices from primary school teachers in China. These researchers developed a large number of materials for improving the methods for assessment. They conducted a study through mixed-research method to analyse the current situation of their educational system. Result demonstrated that teachers improve their quality of teaching and at the same time learners improve their marks. However, teachers were assessment literate and they rejected to develop other theoretical aspects of the field such as historical issues and the current educational policies. They also suggest that more work needs to be done with the development of instruments to track assessment literacy since they need to be more objective towards the educational system and the purpose of

assessment towards learners' needs. Contrastively, Jin (2010) carried out a study to explore the Chinese context in terms of language assessment literacy in higher education. An instrument was administered to 86 Chinese instructors. The instrument was designed into multiple-choice items related to test validity, reliability and other testing principles. Data demonstrated that teachers had a lack of knowledge on the use and application of validity and reliability in language tests. Based on these previous researches, a lack of training in assessment is notorious, but there is also a need to explore what areas are the ones these research context currently require to maximise their teaching and learning efficiency.

Stoynoff (2012) suggested that it is essential to include other research perspectives to analyse the positions of stakeholders in terms of their knowledge, practices and experiences with regards to language testing and assessment. Fulcher (2012) conducted a pilot study to develop materials for language assessment literacy courses. A total of 278 international language teachers took part of this research. They developed an instrument to analyse the essential topics for educational assessment. His contributions towards the field of assessment literacy were to explore the European perspective with regard to previous instruments and he also introduced a model of three levels of understanding assessment literacy that includes knowledge, process and historical and philosophical perspectives of assessment. Lam (2014) analysed the language assessment situation in Hong Kong with teachers from primary and secondary schools. Researcher used an evaluation course sheet, government data, observations and interviews to explore the assessment panorama. He indicated that teachers are not well prepared for their performances as assessors and a training course needs to be developed. Similar study for the basic education, Assunção Tonelli and Quevedo-Camargo (2016) conducted a study to investigate Brazilian teachers of English as an additional language. Data from university students, pre-service teachers, of an undergraduate programme in TESOL completed a survey and interviews to analyse the studies of assessment practices in Brazilian primary schools.

For the purpose of researching assessment for learning or formative assessments against other aspects of language assessment literacy, Riggan and Oláh (2011) analysed the assessment situation in the American State of Philadelphia about the impact of the classroom formative assessment. Data from teachers (n=32) were collected to explore their perspectives and the methods they apply to use formative assessment. Malone (2013) conducted a study based on an online training course. This researcher compared between American language teachers (n=44) and language testers (n=30). The purpose of this research was to compare the contents of the course in an expertise perspective. The results of this study suggest that American language teachers required training in which they can see their progress based on their impact from the training course. In terms of the contents, language testers outlined that there is a need to

## Chapter 2

improve the educational system with formative assessment and validation of educational programmes for language assessment. By the same token, Fulcher (2012) created an online platform with different language testing and assessment resources based on several studies he had conducted throughout the administration of a survey beforehand. Most of the materials that have been developing throughout the last decades had been driven by quantitative research.

Jannati (2015) conducted a study to analyse and examine perceptions and practices related to language assessment from 18 Iranian English language teachers. The researcher used semi-structured interviews to explore their perceptions and practices. Teachers were grouped based on their teaching experience per years. Data shows that teachers were assessment literate, but this did not influence the use of their language assessment practices in their classroom. Berry, Sheenham and Munro (2017) conducted a study to analyse and explore several European language teachers' beliefs, needs and practices with regards to language assessment experiences. They used interviews (n=3), observations (n=3), Group Interview (n=20) and also delivered a workshop (n=48) with topics related to language testing and assessment (e.g., Common European Framework, assessing young learners, assessment for learning, assessment for teachers and assessment resources). Results show that teacher felt more confident after the workshop and they considered that this intervention helped to understand the field of language assessment better. Nevertheless, this workshop did not tackle issues in their participants teaching contexts.

Different educational systems around the globe have been targeted for researchers to explore and examine TALD among different circumstances. Vogt and Tzagari (2014) explored the assessment literacy from language learners and teachers. They developed a survey that included what types of assessment practices European teachers (n=707) frequently use in the classroom and what types of teachers' training needs require to improve the quality of teaching and learning. In addition to this, the conducted semi-structured interviews to deeply analyse the teachers and learners needs. Six European countries took part of this research: United Kingdom, Germany, Cyprus, Italy, Poland and Greece. Results showed that teachers did not have clear thoughts about language testing and assessment, neither theoretical nor practical concepts. Data also reported on that there is a lack of interests in theoretical features of testing such as historical testing and measurement theory. Teachers suggested that training in areas of assessment of productive and integrated skills were more essential than understanding the development of tests. Apart from this, the interview data showed that there is a lack of teachers training regarding language testing and assessment across Europe.

More specifically, Rodriguez and Zabala (2016) conducted a study to examine the higher educational context in Spain in relation to test developers and to construct tests for educational

purposes. They studied a variety of Spanish universities and centres for languages to explore and examine their contents to facilitate mobility using high-stakes tests. This study had a significant impact on teachers' preparation programmes for testing and assessment practices. Another research study conducted to explore Russian secondary teachers of German were surveyed to examine their assessment practices and professional needs. Drackert and Stadler (2018) examined the secondary school teacher' assessment needs for classroom practices and professional development. This study had a policy maker implication due to the recognition of teachers' lack of training in assessment and a re-evaluation of the Basic educational standards for foreign languages was proposed so as to improve the learning and teaching practices through assessment and evaluations. Leong (2014) conducted a study out of the field of language education but in general education. He found out that teachers can amend their teaching practice based on the context. It is worth noting that learners can find several methods of language assessment that might not contribute to or enhance their learning. Regardless of the importance of context, teachers showed that they were assessment literate since they used sophisticated methods of teaching throughout classroom observations. Including a diversity in research methods to observed and track teachers' language assessment literacy levels allow studies to benefit from more evidence to tackle current issues in assessment among research contexts.

Mendoza and Arandia (2009) explored the English language-teaching context in Colombia with regards to language assessment literacy and professional development. The researchers developed an instrument with the most general assessment related topics such as: history of language testing, formative assessment, the Common European Framework (CEF), teachers' needs in assessment, etc. Data suggests that Colombian teachers need more training to develop their lessons and to administer their time between formative and summative. This research study provided a great perspective to international researchers to continue working on the current, but vital features which teachers require to empower their assessment knowledge and skills that they can be reflected in their classrooms. This research indicated that 1) Colombian teachers cannot identify the barrier between formative and summative assessment, and 2) they found that assessment is not adequate for the educational level and the contents of the subject.

Towards the Mexican education, Ruiz-Esparza, Verdugo and Castillo (2015) analysed the feedback that was given to Mexican students (n=8) after their teachers had marked their language tests. They analysed the video recordings made in the classroom and they found out that the assessment practices for delivering feedback changes based on the learners' performances outcomes. This study contributes to the methods of delivering feedback and the assessment practices to cover this task. Feedback has been one aspect that has taken into consideration from Mexican researchers (Martinez-Rizo & Blanco, 2010; Gonzalez, 2018; Roux, 2018). Martinez,

## Chapter 2

Killan, Delgado and Brande (2012) conducted a study to explore teachers' professional development in a south-east region in Mexico. Although this research study was not linked to language testing or assessment, it provided a comprehensive panorama to develop further research in relation to professional development for Basic Education and other educational levels. The authors conducted a study in the south-east part of the state with EFL secondary teachers (n=120) to scrutinise their teaching and professional needs.

This review of the literature indicates that language teachers tend to have less training in assessment or the training that they have received from continuing professional development is not enough. As a consequence, exploring language assessment literacy among educational sectors could provide a coherent aspect on how Tamaulipas EFL teachers consider important or less important in aspects related to the use, administration, marking, communication and identification of language tests and other assessments. In addition, how tests and assessment are constructed could also highlight vital considerations for the research.

### 2.7 Summary

This Chapter presented a comprehensive review of the literature regarding the field of language assessment literacy, the Mexican and Tamaulipas educational backgrounds. This started with the importance of the definitions of the concepts of test, assessment, measurement and evaluation. In addition, terms of assessment and literacy to understand the foundations of language assessment literacy. These are vital aspects to understand their properties and the function they have in the field of educational assessment. Teachers sometimes misunderstand these connotations and that is why it is essential to investigate if teachers know these properties and their purposes. Developing language tests could be seen as a complex activity which requires a lot of attention and systematic steps to properly achieve the purpose of the test.

Language assessment literacy involves knowledge, technical considerations and school-based assessments in which teachers and students are constantly developing to demonstrate that learners have acquire the essential knowledge and teachers have delivered an adequate instruction (Stiggins, 2002). Mexico is a country that follows a pattern of a testing culture which deals with the development of tests and their administration due to the large diversity from National exams (Inbar-Lourie, 2013). The Mexican and Tamaulipas classrooms are places where teachers and students constantly are constantly involved in the activity of assessment (INEE, 2005). There is a lack of studies in regarding the assessment literacy from Mexican teachers of English. Therefore, the first research question stated in this research study is to investigate the status quo of EFL teachers in Tamaulipas (*What is the status quo of EFL teachers' language assessment literacy in Tamaulipas?*) which aims at exploring the development of institutional

tests, the test responsibilities, occurrence of the tests and what language areas are mostly assessed in this Mexican State.

Since language assessment literacy involves three main components: there are eleven standards that were chosen from the literature would explore Tamaulipas teachers' assessment skills and practices. These standards are applied linguistics, educational assessment, institutional assessment policies, testing principles, assessment purposes, assessment processes, educational assessment, language assessment, assessment for learning, communication of assessments and assessment as educational tool (AFT, 1990; Xu and Brown, 2016; Giraldo, 2018). The skills and practices are essential to become a language assessment literate; however, not all teachers require all these standards. There is no evidence of a research conducted in Tamaulipas that allow researchers and teachers to investigate this. In addition, Mexico has two educational sectors: the private and the public. Even though these have different methods of learning, instruction and assessment processes, exploring teachers' language assessment literacy would benefit to identify differences in the field of the English language among the sectors. Also, one more sector emerged from the literature; language institutes also play a fundamental role in understanding teachers' backgrounds in language assessment. Therefore, the second research question is to investigate the differences of assessment skills and practices among educational sectors (*What are the discrepancies, if any, in EFL teachers' language assessment skills and practices among different educational sectors in Tamaulipas?*).

Professional development has also contributed to the field of language assessment literacy where teachers can register to continuing professional development opportunities to familiarize with the backgrounds and learn new skills to be implemented in teachers' classrooms. There is little evidence if Tamaulipas teachers of English have taken any training in assessment. Hence, the third research question for this study would focus on the exploring the teachers' backgrounds in relation to professional development and language assessment (*What are EFL teachers' and programme managers' backgrounds in their professional development towards language assessment literacy?*)

In addition, there is no evidence that I could find from of a workshop based on teachers' current needs for Mexico in terms of a course that covers realistic features on language assessment literacy; as a consequence, the development of a workshop that deal with current national and international perspectives might allow teachers to be more engaged in the field of language testing and assessment (Tzagari, 2009). The four research question deals with exploring the assessment areas which EFL teachers empowered based on the assistance of a workshop (*In what ways has the workshop influenced EFL teachers' awareness of language assessment*

## Chapter 2

*literacy?)* Finally, developing a workshop for this type of purposes is always important to have in mind if this had any impact on the participants in other areas. The fifth research question focuses to measure the consequential validity of a workshop to identify areas where participants apply the knowledge and skills learnt (*What was the impact of a workshop in language assessment literacy development on EFL teachers in Tamaulipas and their educational sectors?*). The following Chapter discusses the methodological procedures for answering these previous research questions. Also, it includes the number of participants, data collection, analysis and ethical considerations.

## Chapter 3 Methodology

This Chapter has been divided into seven sections. The first section (3.1) introduces the two research methodologies and the research design of this investigation. Section 3.2 describes the research context of the study by detailing the location of the study, participants' educational sectors and institutions. The following section (3.3) presents the participants' profiles and samples from each of the data collection phases. Section 3.4 discusses the ethical considerations and how participants' personal background information was securely stored. The development of the research instruments is presented in section 3.5 describing the four phases of data collection. The data analysis methods are presented in section 3.6 and it explains how the data was categorised and reported. Lastly, this Chapter concludes with a summary (Section 3.7) of the most relevant information, as well as includes a table with the research study overview to understand the subsequent chapters.

### 3.1 Research Methodology

Cohen, Manion and Morison (2011) claimed that by identifying adequate research methods, researchers could align the purpose of a research study to obtain substantial evidence for the problem being investigated. Three main crucial aspects guided me (researcher of this study) to make these methodological decisions: 1) to explore and analyse the status quo of language assessment literacy in EFL teachers in Tamaulipas, 2) to examine how teachers' trajectories (e.g., academic, practicum and professional) influence their language assessment literacy among different educational sectors, and 3) to examine teachers' understanding and application of language assessment through a workshop as a professional development opportunity with the current Tamaulipas language educational trends. This research study followed two research methodologies to meet the intended research aims for this investigation: exploratory and experimental methods.

The exploratory research methods aim at observing and studying a phenomenon that occurs in a certain context and brings a comprehensible understanding of a current situation (Saunders, Lewis, & Thornhill, 2012). This method seems to be applicable to research in language assessment literacy in EFL teachers in the State of Tamaulipas to explore three different sectors. This research could potentially demonstrate the strengths and weaknesses of an aspect being studied. In this case, this would demonstrate how language tests and assessment practices are being conducted among teachers within their educational sectors. The exploratory research purpose of this study was chosen because it was vital to understand the current situation (*status*

*quo*) of language assessment literacy in the Mexican State of Tamaulipas. For this research, the 'status quo' refers to the current language teachers' assessments aspects such as: a) test development and responsibilities, b) test relevance and occurrences, c) assessment of language skills and linguistic components, d) assessment knowledge, technical considerations and school-based assessments, e) assessment experiences, f) professional development and g) critical assessment issues in teachers' institutions. Aside from this, participants' previous assessment experiences, trajectories and their institutional contexts were also examined to comprehend the current situation of language assessment. The rationale behind this was to recognise the status quo of EFL teachers and how tests and assessments are perceived and conducted in the classrooms among three educational sectors: the public sector, the private sector and language institutes

This exploratory research study addressed three out of the five research questions: the first research question (*What is the status quo of EFL teachers' language assessment literacy in Tamaulipas?*) deals with the above mention aspects (e.g., test responsibilities, assessment of language skills and linguistic components, test relevance and occurrences, etc.). The second research question (*What are the differences, if any, in EFL teachers' language assessment skills and practices among different educational sectors in Tamaulipas?*) addresses differences among the sectors concerning language teachers and language programme managers. Finally, the third research question (*What are EFL teachers' and programme managers' backgrounds in their professional development towards language assessment literacy?*) examines their academic and professional practices towards language assessment literacy and professional development.

Therefore, a mix-methods approach was applied from the different sources of data which were collected from the survey (Phase 1) and interviews (Phase 2) conducted with language teachers and language programme managers. Mixed-methods approaches use two types of research: quantitative and qualitative. The use of quantitative can provide superficial perspective of a problem, but if a researcher examines in depth could potentially provide a coherent analysis where both methodologies could provide a better panorama of the current situation (Hernandez, Fernandez & Baptista, 2003). Qualitative and quantitative approaches had equal importance to answer these three research questions. The survey and interview findings were individually analysed and then scrutinise to highlight the similarities and differences among the participants, as well as their educational sectors. Although some researchers have used and recommended classroom observations methods to observe and examine how teachers conduct tests and other assessment skills and practices in their classroom (Scarino, 2013, O'Sullivan, 2017); however, this study did not include this method due mainly to a lack of permissions from the educational institutions and restrictions of relating to time and its geographical nature. Therefore, I decided to

include language programme managers to support teachers' perceptions as they could provide more evidence to their educational sectors.

Tolmie, Muijs and McAteer (2011) suggest that exploratory research may bring benefits to observe and identify the surface of a phenomenon, but this could potentially not provide clear conclusions about the problem being studied. As a result, a follow-up study was conducted. Experimental research design enables researchers to randomise participants to examine the fruitfulness of an experiment. Hence, participants from the three educational sectors would be used to examine in-depth their language assessment literacy. Cohen, Manion and Morison (2011) provide relevant features to highlight how to conduct experimental research: the number of samples, the variables to explore and the problems in the validation process. The experiment, in this research, consists of developing a workshop that would be created from the current teachers' needs about participants' assessment skills and practices.

Quasi-experimental research was included based on two methodological decisions: 1) to measure and observe the most and least progress and 2) the impact of the workshop on EFL teachers in different areas of their practicum. This experimental research relates to the last two research questions of this research study; the fourth research question (*In what ways has the workshop influenced EFL teachers' awareness of language assessment literacy?*) explores the most and least significant areas where EFL teachers consider essential and made more and less progress during the workshop. The fifth and last research question (*What was the impact of a workshop of language assessment literacy development on EFL Teachers in Tamaulipas educational sectors?*) examines the degree of impact towards their professional, educational, and institutional aspects. These two research questions specifically dealt with the instruction of a workshop in language testing and assessment developed from the findings of the research questions 1, 2 and 3 (exploratory research). The experimental approach was designed to *develop* and *measure* the progress, as well as the potential impact from the workshop on EFL teachers in Tamaulipas.

A concurrent-mixed method approach was applied to analyse these data (experimental research). This methodology has also been conducted in studies from Creswell (2003) and Mertler (2009), in which data were analysed simultaneously and complemented by the different data sources. These researchers had used tests, interviews, and surveys in their research studies to identify differences and expand the researchers' view in relation to a target domain. Having different sources for data collection could broaden the perspective to understand certain phenomena around the environment of the research (Savin-Baden & Howell Major, 2013). Participants' materials and tasks from the workshop (Phase 3) were analysed simultaneously. A

focus group interview (Phase 4) was also conducted to evaluate the efficiency of this workshop and the implications towards the use of professional development to improve educational assessments among different educational sectors.

The following table (see Table 13) describes the research design and the data collection phases. On the left side of this table, it presents the four data collection phases and the pilot studies conducted to examine the validity of the instruments. On the right side, the data collection phases are reported with the date when they were conducted.

**Table 13 Research Design, Data Collection Phases and Chronological Description**

Research Design	Data Collection Phase
Pilot of the Survey and Interviews	<u>Conducted:</u> March 2017 (Quantitative) April 2017 (qualitative)
Final Administration of the Survey	<b>Phase 1</b> <i>Teachers Assessment Literacy Development Inventory (TALDI)</i> March 2018- September 2018
Final Administration of the Interviews	<b>Phase 2</b> Semi-structured interviews with EFL teachers and programme managers <i>April 2017 / September 2018</i>
Pilot of the workshop	<u>Conducted:</u> August 2018 (Online version with 4 Mexican participants)
Instruction of the workshop	<b>Phase 3: Instruction</b> Tamaulipas Teachers Assessment Literacy Development Workshop <i>November 2018</i>
Focus Group to measure the progress and impact of the workshop for EFL teachers in Tamaulipas	<b>Phase 4: Evaluation</b> Focus Group (Skype Call) February 2019

An overview of this investigation with the research questions, research aims, and methods of data collection and analysis can be found at the end of this Chapter (see Table 24). After explaining the research design and how the research questions were answered, it is essential to comprehend the background of this research context. The following section describes the research context: location of the study, institutions and the sectors that were explored.

### 3.2 State of Tamaulipas: Institutions and Educational Sectors

This section describes and provides current information about the institutions that took part in this research. Some figures are also provided to understand the educational settings of the State of Tamaulipas about the number of institutions, students, and teachers. This research explored three different educational sectors: the public sector, the private sector and language institutes, where the English language is taught as a foreign language (EFL) in Tamaulipas. The institutions that were taken into consideration were upper-secondary schools, higher education institutions and language institutes. The rationale behind this decision was to unify these three

types of institutions to further explore the three educational sectors. The exploratory and experimental studies were conducted in the North-East Mexican State. This state, Tamaulipas, has been well-known due to the geographically close relationship to the United States of America for the English language learning opportunities. It is worth noting to mention that Tamaulipas was one of the first Mexican States that included English as a foreign language as a subject primarily in the private educational systems in 1926, but it was not until 1993 that English was incorporated into the public sector (Roux, 2018).

The Ministry of Education in Tamaulipas (SET) is responsible for certifying and approving educational programmes in this state from early learning stages up to higher education. According to the most recent statistics published from SET<sup>10</sup>, the number of institutions (e.g., upper-secondary and higher education) in the State of Tamaulipas from the years 2017 to 2018 was a total of 485 upper-secondary schools: 235 public and 250 private. For higher education institutions, there was a total of 193: 61 public and 132 private. There is not an accurate number for language institutions in Tamaulipas, but according to the latest survey from the Mexican Association for English Language Teaching in Tamaulipas (MEXTESOL), it was stated that there were approximately 550 language institutes from 2016-2018 (MEXTESOL: Tamaulipas Chapter, 2018). In relation to the number of students, SET indicated that there were 147,010 students registered from upper-school: 103,251 public and 43,759 private, and for higher education 124,094 students: 79,412 for public and 44,682 for private. There were 8,218 teachers employed in upper-secondary school and 10,694 teachers at higher education institutions. From this statistical report there is a notorious number of students from the public sector (SET, 2019).

The number of hours varies based on the type of institution. For example, a public upper-secondary school has approximately 5 hours a week, whereas in a private school the number of hours per week can reach up to 10. For higher education institutions, specifically, the private ones tend to have more hours of English language programmes than the public institutions (British Council, 2015). All these previous institutions include EFL courses as part of their official curricula. Intending to understand the research context of the educational sectors and institutions that were explored and considered for this study, the following section includes a summary of the educational sectors and the institutions.

---

<sup>10</sup> The Ministry of Education in Tamaulipas (SET, Spanish acronym) published online the most recent statistics (2017-2018) in general education. This data can be obtained on: <https://www.tamaulipas.gob.mx/educacion/wp-content/uploads/sites/3/2018/06/anuario-2017-2018.pdf>.

### 3.2.1 Educational Institutions

This section seeks to understand the backgrounds of the schools and organisations that took part in this research study. The schools and educational organisations were the following:

**Upper-Secondary Schools (High-schools)** - Students at this type of institution are typically aged between 15-18 years old. Some institutions provide different programmes; for example, a technician degree or speciality in one area (e.g., Technician in Accounting, Electrical or Clinical studies, etc.) or general bachelorettes. These two programmes are covered with different subjects such as Spanish, Mathematics, History, Chemistry, etc. Several of these institutions also offer second language (e.g., English, French, and Chinese) courses as mandatory. Some private schools require their students to certify their level of English to successfully achieve these academic studies and continue to the next educational stage, which is higher education. No public schools require students to certify their second language proficiency. English is seen as one of the subjects from the whole curriculum, but not as an essential or vital subject.

**Higher Education Institutes**- Students at this type of institution are at least 18 years old and are registered to bachelor degree programmes from different disciplines in which the English language is included as a mandatory subject. Several institutions require students to demonstrate language proficiency level B1 (CEFR) to be able to graduate and be awarded their intended degree (UAT, 2011). Institutions offer English courses differently; for the public institutions, English could last up to one year with 4 hours a week.

**Language Institutes**- Students' ages at language institutes vary since these institutions aim at registering students who desire to learn or reinforce a second language. These institutions are only responsible for teaching languages. These institutions are sometimes incorporated into higher education institutes, but they can work differently and independently from higher educational programmes.

### 3.2.2 Educational Sectors

This section aims to provide a description of the three educational sectors which this research aims to explore and investigate their EFL teachers. This description explores the current settings such as the number of hours for English instruction, classroom sizes, among others. The three sectors explored and analysed were the following:

**Public Sector:** this sector comprises of institutions that are fully or partly funded by the local, state, and federal governments. These institutions follow the National curriculum developed by the Ministry of Education. Students regularly pay lower tuition fees every academic term: annually or per semester. This type of sector has several qualities such as the use of technological

devices and a significant number of students in the classroom. In relation to the English language, learners have a limited number of hours per week. Teachers are hired based on the National standards for education.

**Private Sector**- this sector gets funding from external sources and not directly from the government. Consequently, students need to pay much more to be registered in one of these schools. They develop or share the same curriculum as the public, but the instruction delivery and contents differ, and they have total autonomy from the assessments and students' registrations. Technological facilities are more frequently used in these institutions and the number of students is lower. Languages are considered a fundamental aspect of private institutions, mainly in higher education institutions.

**Language Institute**- this sector only deals with those institutions where languages are taught solely. Their priority is to provide instructions for foreign languages. Their tuitions fees are not sometimes that accessible and students take these courses as an extra-curricular opportunity to expand their second language proficiency trajectories. These institutions also offer courses to prepare for international certifications such as the Cambridge main suite (e.g., A2 Key English Test, etc.), Educational Testing Services such as TOEFL, and other national and international certifications.

Overall, this research study collected data from a total of 20 institutions in Tamaulipas: 8 upper-secondary schools (4 public and 4 private), 7 higher education (3 public and 4 private) and 5 language institutes. The following section discusses the participants' profiles and the samples of each data collection phase.

### 3.3 Participants' Profiles and Samples

Two types of stakeholders were considered for participation for this research project: EFL teachers and language programme managers. Participants were chosen based on their profiles and the educational sector where they were currently working. The key characteristics of the chosen participants and institutions were the following:

- At least 18 years old and working in the State of Tamaulipas, Mexico,
- Working in one of the three main educational sectors being explored (i.e., the public sector, the private sector or language institute),
- Employed as an EFL teacher, language programme manager and/or a test developer from an Upper-secondary school, higher education institution or in a language institute,
- Hold a full-time or part-time position,
- Have access to a language classroom, and

- Willingness and motivation to participate in at least one of the four research phases.

The process of recruiting participants started in the spring of 2017. During this time, I was in Mexico and had the opportunity to discuss with language programme managers, teachers, and other institutional authorities. I was also able to get a database from the local Association for English Language Teachers in Tamaulipas (MEXTESOL), which was used to be able to contact potential participants through different forms of communication such as email, calls or meetings. To invite them as participant, I was required to design a participation letter to the head of each institution; only language programme managers or directors of the institutions received this letter. This letter was written in the potential participants' first language (Spanish) for their institutional authorities to obtain all legal permissions required to conduct this study. An example of a participation letter that was approved by an institution can be found in Appendix C with a translation into English. The sample of the participants varied from each research phase. The following section presents the number of participants in each of the four data collection phases. An in-depth description of these data collection phases is presented in section 3.5.

### **3.3.1 First Phase (Survey)**

The first phase was the administration of the Teachers' Assessment Literacy Development Inventory (TALDI), which was the main source for collecting quantitative data. A total of 300 surveys were printed and delivered to academic coordinators, directors, and language programme managers from different educational institutions where the English language was taught as a foreign language at the Capital city of the State of Tamaulipas, Victoria. Surveys were returned to the researcher to be analysed ( $n=186$ ), but only 163 surveys were valid and taken into consideration for the process of data analysis. Moreover, the researcher also contacted institutions outside the Capital City, to collect more data and be able to contextualise the research across Tamaulipas. Approval letters, consent forms and the surveys were sent electronically to their language programme managers, so they could print and distribute the instruments. Surveys were returned via post. A total of 152 surveys were received from four different cities of the State of Tamaulipas (Rio Bravo, Nuevo Laredo, Matamoros, and Tampico). The intention of taking into consideration participants from other cities was to increase the number of the sample to obtain a distributed number in each of the educational sectors. It is worth noting that I only examined and explored the teachers' backgrounds and language assessment literacy and did not scrutinise participants' students. Also, this survey had previously contemplated as an e-survey to be conducted, however, this was discarded due to a low number of participants and the lack of valid responses.

Overall, 315 surveys were valid from participants who completed all the sections of the instrument. Table 14 summarises the participants' backgrounds in relation to their gender, age, highest academic qualification, teaching experience and previous language testing and assessment training taken from at least 18 months before the administration of TALDI.

**Table 14 Participants from Phase 1 Survey: (TALDI n=315)**

Theme	Reference	Public (n=105)		Private (n=108)		Lang. Institute (n=102)	
<b>(2) Gender</b>	Male	48	45.7%	36	33.3%	31	30.4%
	Female	57	54.3%	72	66.7%	71	69.6%
<b>(3) Age*</b>	18	-	-	-	-	-	-
	18-20	3	2.9%	-	-	6	5.9%
	21-29	43	41%	65	60.2%	36	35.3%
	30-39	43	41%	36	33.3%	53	52%
	40-49	14	13.3%	6	5.6%	7	6.9%
	50-60	2	1.9%	1	.9%	-	-
	>60	-	-	-	-	-	-
<b>(4) Highest academic degree</b>	Tech.	-	-	-	-	-	-
	B.A.	76	72.3%	57	52.8%	71	69.6%
	Dip.	7	6.7%	24	22.2%	18	17.6%
	M.A.	22	21%	27	25%	12	11.8%
	PhD.	-	-	-	-	1	1%
<b>(7) Teaching experience</b>	<1	-	-	6	5.6%	3	2.9%
	1-2	3	2.9%	8	7.4%	3	2.9%
	3-5	40	38.1%	40	37%	26	25.5%
	6-10	45	42.9%	43	39.8%	51	50%
	11-15	15	14.3%	8	7.4%	18	17.6%
	16-20	-	-	2	1.9%	1	1%
	>20	2	1.9%	1	.9%	-	-
<b>Previous training in Assessment</b>	Yes	16	15.2%	28	25.9%	46	45.1%
	No	89	84.8%	80	74.1%	56	54.9%

A total of 200 were female and 115 were male participants. This was made up of 303 EFL teachers (96.2%), 11 language programme managers (3.5%) and 1 participant with an administrative position (0.3%) participated. All participants (n=315) were responsible for teaching EFL in a classroom, even if they have two or more positions in their same workplace. With regards to their educational sector, the public sector had 100 language teachers (95.2%) and 5 programme managers (4.8%). The private sector had 103 language teachers (95.4%), 4 programme managers (3.7%) and 1 administrative position (0.9%). The language institute cohort had 100 language teachers (98%) and 2 programme managers (2%). Therefore, 33.3% from the public sector (n=105), 34.3% the private sector (n=108), and 32.4% participants from language institutes (n=102). Data were collected from five different cities from the State of Tamaulipas: 17.5% from Tampico (n=55), 13.6% from Matamoros (n=43), 9.6% from Rio Bravo (n=30), 7.6% from Nuevo Laredo (n=24), and the majority (51.7%) from the capital city of the state, Victoria, with 163 participants.

### 3.3.2 Second Phase (Interviews)

The second phase consisted of conducting semi-structured interviews. The selection of the participants was based on their educational sectors, including those teachers who were working from different educational levels (e.g., upper-secondary school, higher education institutions and language institutes). A total of 9 participants were invited throughout emails since they provided this information while completing TALDI and the other three from the institutions where these interviews were conducted. The rest of the interviewees (n=3) were conducted at the stage of the pilot study in March 2017. IN addition to these participants, one language programme manager from each sector took part in this phase. The interviews took place in five institutions in Tamaulipas. Interviews were conducted in both languages English (8) and Spanish (7) as some participants felt more comfortable speaking their first language. The interviews were audio-recorded with each one lasting approximately from 15 to 20 minutes.

Table 15 presents the interviewees' personal information including the participants' ID code ("PUB" as public, "PRIV" as private and "LI" as language institute), age, gender, highest academic and teaching qualifications (e.g., Bachelor's degree (B.A), Master's degree (M.A), Diploma (Dip); teaching qualifications such as: Teaching Knowledge Test (TKT), In-Service Certificate of English Language Teaching (ICELT)), previous training in language testing and assessment (LTA) and participants' teaching experience (T.E.) reported in years. It is worth noting that Chapter 5 presents the findings of the interviews, the extracts of the participants were signposted as (INT).

**Table 15 Participants from Phase 2 (Interviews)**

Participant ID Code	Age	Gender	Qualifications	LTA	T. E.
PUB-T1	32	Male	MA/TKT	X	10
PUB-T2	51	Female	BA/ICELT	X	23
PUB-T3	42	Female	Dip/ICELT	Y	21
PUB-T4	25	Male	BA/TKT	X	5
PRIV-T1	36	Female	BA/ICELT	X	9
PRIV-T2	24	Female	BA/TKT	X	2
PRIV-T3	29	Female	Dip/TESOL	Y	6
PRIV-T4	57	Female	BA/TKT	X	26
LI-T1	29	Male	BA/TKT	X	3
LI-T2	43	Male	MA/TKT	X	21
LI-T3	34	Female	BA	X	12
LI-T4*	30	Female	MA/ICELT	Y	6
PUB-LPM	47	Male	MA/ICELT	Y	12
PRIV-LPM	32	Female	BA/TKT	Y	5
LI-LPM	38	Male	BA/TKT/ICELT	Y	7

\*Note: Transcription included in Appendix F.

A total of 15 participants took part in the semi-structured interviews consisting of 12 EFL teachers (4 from the public, 4 private and 4 language institute), as well as 3 programme

managers, one from each educational sector. The number of teachers was decided based on the type of institution (2 from upper-secondary and 2 from higher education institutions). In terms of participant's gender, 9 were females and 6 males. Participants' age range was from 24 to 57 years old. Four participants had been awarded Masters' Degrees in TESOL, Innovations in Education, Applied Linguistics and Business and Management; whereas the remainder held Bachelors' Degrees in Language Educational fields (e.g., Applied Linguistics, Language Pedagogy, etc...), but also others disciplines (e. g., Tourism, Marketing and International Business). Two participants indicated that their higher degree achieved was a Specialisation or Diploma in Mathematics and one from English Language Teaching. International Teaching Qualifications were also reported among the participants: Teaching Knowledge Test (n=8), In-service Certificate of English Language Teaching (n=6), and TESOL Certificate (n=1). Participants' teaching experience was from 2 to 23 years. Participants also indicated that only 6 out of the 15, including the three programme managers, had taken a course or training in language testing and assessment within 18 months from the moment of conducting the interviews.

### **3.3.3 Third Phase (Instruction of the Workshop)**

A total of 22 participants were interested and confirmed taking part in the workshop two weeks prior it started in November 2018. From those 22 participants, 19 attended the first lesson, but at the end of the first week, only 18 participants remained and completed the tasks required. However, by the second week, 15 participants remained registered, but only 10 fully completing the course contents and tasks. For this study, I decided to only explore these 10 participants who completed all tasks including the pre- and post- measures, practical and reflective tasks, and the course evaluation sheet. EFL teachers and programme managers were invited, but none of the language programme managers was able to attend due to time restrictions and workload. Also, I was interested in the reasons why participants did not continue with the workshop.

Table 16 presents the participants from this workshop. Seven participants were female and three were male. They were divided into clusters in relation to their educational sectors. Participants indicated that they had not taken any previous training course in language testing and assessment in the last 18 months with the exception of one participant indicated that had taken a course before. The participants' academic backgrounds showed seven participants held bachelor's degrees and three masters' degrees. Eight out 10 have achieved an initial qualification (TKT and CELTA) and only one participant held an in-service teacher certificate (ICELT). The participants' codes are not the same as the participants from the semi-structured interviews (Phase 2).

**Table 16 Participants from Phase 3 (Workshop)**

Participant ID	Age	Gender	Qualifications	LTA	TE
PUB-T1	28	Female	BA/TKT	X	2
PUB-T2	30	Female	BA/TKT	X	4
PUB-T3	24	Male	BA/ICELT	X	2
PUB-T4	38	Female	MA/TESOL	X	7
PRIV-T1	29	Female	MA/CELTA	X	3
PRIV-T2	32	Male	MA/TKT	Y	8
PRIV-T3	23	Female	BA/TKT	X	2
LI-T1	35	Male	BA/ICELT	X	9
LI-T2	25	Female	BA/ICELT	X	2
LI-T3	31	Female	BA/ICELT	X	3

### 3.3.4 Fourth Phase (Focus Group Interviews)

The purpose of the fourth phase was to seek further data to be able to triangulate the different phases from this research. This last phase of data collection was to evaluate the participants most and least interest in the subject of language testing and assessment and to measure the impact of the workshop on participants. All participants from Phase 3 (workshop) were informed about a focus group interview through a video-call would happen weeks after the workshop. Participants were invited through emails to take part, which was conducted with the assistance of Skype in February 2019.

In this phase, a total of six participants, 2 from each cohort (e.g., public, private and language institutes), took part in this focus group interview. No LPMs were considered since there was no participation from them in the workshop. The audio was recorded and lasted approximately 65 minutes. Participants had the opportunity to teach in their classroom for a period of 14 weeks after the instruction of the workshop. These participants fully completed the contents of the workshop syllabus, tasks and were asked questions to measure their progress and impact of the course on these six participants. More information about what participants discussed can be found in Chapter 6. Table 17 presents the participants' personal information such as the same participants' ID from the workshop, age, gender, previous qualifications, and teachers' experiences (T.E.) reported in years.

**Table 17 Participants from Phase 4 (Focus Group)**

ID Code	Age	Gender	Qualifications	T.E.
PUB1	28	Female	BA/TKT	2
PUB2	30	Female	BA/TKT	4
PRIV1	29	Female	MA/CELTA	3
PRIV2	32	Male	MA/TKT	8
LI2	23	Female	BA/TKT	2
LI3	31	Female	BA/TKT	3

Overall, the exploratory research collected data from Phase 1 (Survey: TALDI) with a total of 315 participants and Phase 2 (Semi-structured Interviews) with 15 participants, whereas the experimental study had a total of 10 participants in Phase 3 (Workshop) and six from the Phase 4 (Focus group). The following section describes the ethical considerations that this research took into consideration towards participants' personal information.

### **3.4 Research Ethical Consideration**

As part of this research to collect and protect participants' background information data, ethical approval was indispensable in order to meet the required considerations from the University of Southampton Data Protection Act. 1998. This university finalised the evaluation of the research protocol and plausible risks by March 2017. Therefore, I obtained the Ethical ID authorisation number (25427). All participants, regardless of their participation from the data collection phases, had to demonstrate their willingness to be part of this research by signing a consent form and a participant information sheet. Participants were informed by me that no personal information would be reported in the findings nor further publications from this research study.

Samples of these documents could be found in Appendix A (Consent Form) and Appendix B (Participant Information Sheet). There were no reported risks during the process of conducting the instruments and all the participants agreed to participate. A total of 368 consent forms were signed from the surveys, interviews, materials from the workshop and focus groups have been kept at the researcher's home in Mexico. The audio recordings from the semi-structured and focus group interviews were stored at the researchers' audio-recording device and personal computer. The following section of this chapter describes how the research instruments were developed, piloted, and administered.

### **3.5 Development of the Research Instruments and Interviews**

This sub-section includes the development, pilot, and final administration of the instruments. As has been discussed before, four phases were followed to gather enough information to answer the five research questions from this study. This sub-section is divided into four parts. Section 3.5.1 discusses the survey entitled Teachers' Language Assessment Literacy Development Inventory (TALDI) which includes the procedures of development of the two versions, the piloted one and its final administration. The Section 3.5.2 the semi-structured interviews are described with a rationale for this method, the purpose and focus. Section 3.5.3 presents the considerations and procedures to develop the workshop syllabus, pre- and post-measures, tasks, and an evaluation sheet. Besides, a description from the three stages of the

workshop: pre-workshop, during the workshop and post-workshop. Finally, Section 3.5.4 discusses the focus group interview with the rationale, purpose, and methods for conducting this phase.

### **3.5.1 Teachers' Language Assessment Literacy Development Inventory (TALDI)**

This quantitative survey named as Teachers' Assessment Literacy Development Inventory (TALDI) had been termed with two keywords that allowed me and the participants to highlight its purpose. On one hand, the word '*development*' refers to the importance of teachers' continuing professional development (CPD) which included aspects to enhance their assessment' skills and practices, as well as preferred methods for learning to collect evidence for the development of Phase 3 (workshop). This instrument collected evidence of professional development opportunities (e.g., courses, workshops, academic trajectories etc.) which teachers could have taken in the past (for at least 18 months before) to boost their professional development in language testing and assessment. On the other hand, the term '*inventory*' deals with a set of different technical aspects in the instrument (e.g., knowledge of assessment, technical considerations and school-based assessments) to analyse specific features related to language testing and assessment, at the teacher level and institutional level among educational sectors where assessment takes place. The last version of this instrument can be found in Appendix D.

The following sub-sections are divided into three parts. The first section deals with the survey development (TALDI) (Section 3.5.1.1), then, the pilot study 1 (Section 3.5.1.2) to validate the instrument based on recommendations that were made from participants and statistical tests to ensure the final administration for this instrument, and the third section (Section 3.5.1.3) deals with the final version of TALDI to highlight how the instrument was comprised and administered.

#### **3.5.1.1 Survey Development**

The use of surveys in educational research has gained a lot of attention due to its complexity in developing and collecting adequate data. Cohen, Manion and Morrison (2011) indicate that surveys aimed at gathering large-scale data and they can be oriented to certain particularities such as participants' opinions, conditions and rating (2011: 128). The development of this survey was a challenging process due to the technical jargon used and the scope of the research since the main purpose of this research tool was to examine and analyse the status quo of EFL teachers' language assessment literacy in relation to their assessment' skills and practices in the classroom. Surveys should underpin its purpose and content to be valid and reliable; as well as to be planned efficiently with guidance and techniques to have a flow sequence for the

participants. For the purpose of this survey, the development consisted in identifying key elements as its contents, purpose, population, analysis and interpretation.

There were two versions of TALDI; the preliminary draft conducted during the pilot study (March 2017), and the final version administered during the months of September 2017 to September 2018. The pilot version of the instrument was initially divided into three sections: a) *participants' personal backgrounds*, b) *professional development in language testing and assessment and professional development opportunities*, and c) *language assessment skills and practices*. The following description presents an overview of these three parts included in this piloted instrument. It is worth noting that a total of 100 items were developed for this piloted version of TALDI.

### **Participants' Personal Background (Piloted version)**

There was a total of 18 items that aimed at exploring the participants' personal background information such as educational sector, gender, age, teaching experience and academic trajectories. It also investigated the background of the institution in terms of the number of hours teachers spent in their classroom as well as the number of students in each. This section was essential in order to classify the participants from different cohorts and to explore their educational sectors. This had a total of 2 items that facilitated the purpose of clustering the participants into the three main educational sectors and classifying the institutional operational system of assessment in Tamaulipas.

### **Professional Development and Language Testing and Assessment (Piloted version)**

The second section of the instrument collected participants' teaching practice, professional development and needs for assessment throughout items questioning previous educational courses or academic programmes taken by the participants. This section was vital to understand the background of professional development from the participants. There were three items with 32 general topics related to language testing and assessment (theory and practice) in which the participants needed to categorise based on their confidence and professional needs, as well as the level of their understanding in certain concepts of testing and assessment. There was one item (36) that allowed participants to self-categorise topics related to assessment from '*less essential to very essential*' with the intention to develop the skeleton for the workshop. Also, there were three items (37–39) that tackled issues related to the workshop, it asked participants about their interests in taking part in a professional development opportunity free of charge and also enquired why participants would not attend to this opportunity should that was the case.

### **Institutional Assessment Skills and Practices (Piloted Version)**

This section used a previous research instrument that had been adapted from the original Assessment Practices Inventory (API) by Zhang and Burry-Stock (1997, 2003). This instrument followed the 7 standards from the American Federation of Teachers (AFT, 1990). This 1997 version from the instrument included 67 items which were designed into a double- multiple Likert scales. It dealt with two aspects; the self-perception on how teachers considered themselves competent in terms of the skills required and the frequency of assessment practices. The rationale for selecting this instrument was because it obtained a high-reliability level ( $r = 0.97$ ). It also collected two essential aspects for the purpose of this research skills and practices under the umbrella of the AFT standards of Assessments. From Zhang's and Burry-Stock's version several items were removed, specifically those that were not included in the scope of the research project. An example of these items was related to basic education contexts which included the communication of assessments to parents or guardians. The piloted version was adapted with a total of 61 items divided into 8 standards, including the standard for Assessment for Learning.

#### **3.5.1.2 Pilot Study 1**

A small study, before conducting the main one, may provide relevant information about the design, administration and analysis of the research instruments. Therefore, a pilot study usually is considered vital, especially for those research instruments which had been adapted or newly developed (Cohen, Manion and Morrison (2011)). A pilot study also tends to guide the researcher to prevent issues related to the study and maximise its efficiency by reducing time, stress and anxiety. Dörnyei (2007) suggests that a pilot study should be conducted mainly for quantitative tools as these require detailed attention for its purpose and to avoid any bias. By conducting a small empirical investigation, it allowed me to measure the feasibility of the project and how to deal with research inaccuracies. In addition, presenting some preliminary results might bring a significant number of advantages to revise and validate the research instruments and to be aware of further amendments.

A pilot study was carried out with EFL teachers and programme managers ( $n=77$ ) during March 2017. TALDI was administrated with the intention to: a) explore preliminary findings in relation to participants' self- perceived assessment skills and practices (TLASP), b) investigate their professional development opportunities, c) develop a preliminary skeleton and potential materials for the workshop, d) identify the main sources in which how participants had had access to teacher training before, and e) to obtain feedback on any improvements and recommendations from the participants. Nevertheless, the vital reason behind conducting this pilot study was to

collect as much feedback possible to improve this research tool to avoid complex language and technicalities related to the field.

The instrument was conducted in two different settings. Participants were invited to attend a meeting in which the researcher explained the purpose of the research and clarify how to complete the instrument successfully. Ethical considerations such as participants' consents and participant information sheets were discussed, distributed and signed. Several participants completed the questionnaire during this meeting since they had several queries with regards to the structure or technicalities and to avoid any confusion whilst completing the questionnaire. Some questionnaires were given to those teachers who were unable to complete the instrument due to the lack of time during this meeting. TALDI was administrated as a paper-based instrument. The outcomes of this pilot study were beneficial due to flagging up of several problematic items in TALDI. Participants mainly mentioned that the length of the instrument was "longer than usual" and it was favourable to decrease the number of items. I received some written feedback from some misspelling and the technical jargon used that participants who were unable to understand several items.

In relation to the research study aims and structure of the instrument, there was a need to modify and to divide the instrument from three sections into five sections after the pilot study. In terms of TALDI function and form of the survey, the pilot study 1 aided at diving professional development from language assessment and included one more section with a qualitative aspect to collect more evidence regarding participants' institutional assessment practices, contexts, sectors and experiences, but also to delete several items that were ambiguous and repetitive.

In the first section, participants' personal background, items such as language learning and teaching qualifications (10 and 11) were redundant and not needed in this section; they were moved to professional development instead. Also, items 13 and 14 (Mexican geographical regions) should be excluded due to the scope of the study, after a critical evaluation on the feasibility of the study, Tamaulipas was only considered.

In the second section, the researcher had to divide the items because they were confusing for the participants and these were not as clear between professional development and in the field of language testing and assessment. The first one should deal with the professional development in broadening aspects in order to identify participants' levels of interest towards assessment among other elements of professional development. The second solely in participants' self-perceptions towards their own levels of confidence and interests in testing and assessment.

The third section from the piloted version obtained more feedback from the participants. This one was in which participants struggled with the most and this is the fundamental piece of the instrument to meet the research objectives. The double-multiple choice items were complicated for some participants due to the long section and they had to tick two responses for each item. This section needs to categorise the items based on the standards of the language assessment literacy in the classroom which were intended to examine for this research. The instrument from Zhang and Burry-Stock (1997) did not evaluate knowledge: in fact, it collected participants' skills in different areas answering the question (how skilled are you in...?). Therefore, I needed to adhere to some elements that were more theoretically linked to the use of language assessment for enhancing the knowledge of assessment.

A new section was created after conducting the pilot study with the intention to gather more evidence about the institutional assessment practices and the accessibility to the development of tests at participants' educational sectors. This was flagged up because not all teachers developed their language tests. A selection of open-ended questions should bring a qualitative perspective to the instrument, I must mention that the use of these items could provide more information if these are analysed qualitatively. Moreover, qualitative data in this research needs to be taken into consideration because they could be beneficial to underpin certain aspects that were not considered at the time of the development of the survey (Dörnyei, 2007). Overall, after an in-depth revision of the instrument (with participants' recommendations and statistical evidence in relation to understanding the data collected) showed that from a total of 100 items only 88 remained for the purpose of this research study, and this resulted in time of completion from a maximum of 40 minutes reduced up to from 25 to 30 minutes as maximum.

At the statistical level, Cohen, Manion and Morrison (2011) highlighted that validity and reliability of quantitative instruments can be obtained by the reliability Index. These researchers indicate that an instrument with an average of .70 based on the Cronbach's Alpha index of reliability usually is a reliable instrument. According to SPSS findings, TALDI (piloted version) obtained a Cronbach's Alpha of 0.74, which could be interpreted as a reliable instrument. The following section presents how the final version of TALDI was developed per each section.

### **3.5.1.3 Final Version of TALDI**

The final version of TALDI was conducted during the months of September 2017 to September 2018. This final version of the instrument comprises five sections: a) participants' personal information, b) participants' experiences and educational contexts, c) professional development in language testing, d) continuing professional development, and e) practitioners' assessment skills and practices. An overview of the final instrument and the rationale of why I

decided to input certain items is presented below by describing each section of the final version of TALDI.

### **Section 1: Participants' Personal Information Backgrounds**

A total of 11 items were remained in this section. The list of items of this section was to identify different aspects of the participants such as the type of institutional context, gender, educational field, and levels of English language teaching. The most relevant aspect of this section was educational background (including the three sectors being explored for the purpose of the study), the highest educational programme awarded (adding a new element which is the "Diploma Course" which for Mexican Educational standards is known as "Specialisation"), teaching experience, working status as (full-time or part-time), and classroom characteristics: teaching hours and the number of students. After the pilot study, this section was barely modified to present and collect information easily for the researcher.

### **Section 2: Teachers' Experiences and Educational Context**

This section was added after the pilot study since there was a lack of information related to participants' institutional assessment policies and practices. I decided to include seven open-ended items (items 12-18) since this was a method in which participants could expand their responses and obtained an in-depth analysis for the purpose of the study. The first question dealt with the methods in which participants assess their students (*How do you assess your language students?*), the second question aims at collecting evidences who develops testing materials for formative and summative assessments (*Do you design the tests that are administered in your classroom? if not, who develops the tests?*), then, the following was related to other methods of assessment, especially alternative assessments (*Apart from language tests, what else do you use to assess your students?*), the fourth was on how teachers select, adapt or develop language tests and materials (*How do you choose your methods of assessment?*), the fifth dealt with the aspects of identification about the usefulness of the test instruments teachers use for their students (*How do you realise that the methods for testing and assessing your students are the most adequate?*), the sixth regards to the regulations and assessment policies, as well as code for ethics in testing and assessment (*Are there any code of practices for development, administration, marking or communication of assessments in your institution?*). Lastly, the last item aimed at examining critical issues that participants have dealt with the use of assessments (*Have you identified any problems with the use of assessment in your institutions?*) Since this type of responses cannot be measured quantitatively, a qualitative analysis needs to be conducted.

### **Section 3. Professional development in Language Testing**

This section has 5 items (19-23). The first four items dealt with participants' interests in language testing and assessment, teachers' confidence with language testing, as well as, classroom assessment, and participants' previous training in this field. This section also included one more item that dealt with the participants' interests to register in a free of charge workshop and gathered information about the most relevant topics of language testing and assessment. This item with 26 assessment themes (e.g., History of language testing, development of language tests, using holistic or analytical rubrics, etc.) in which participants required to choose from "1=not essential to 5= very essential", based on their interest in participating in a language testing and assessment workshop. This list was essential to underpin the contents of the workshop for phase 3.

### **Section 4. Continuing Professional Development**

This section includes 11 items (questions 24-34) to examine and classify how EFL teachers in Tamaulipas have developed their professionalism throughout Continuing Professional Development (CPD) opportunities to enhance their teaching practice in other professional practices such as planning, classroom management, etc. This section was divided into 3 parts. The first part collected participants' responses with respect to their perceptions of CPD within their working institution (items 24 to 26). Items 27 and 28 collected participants' evidence for language proficiency and teaching qualifications.

The second part of this instrument, participants were required to consider which components of professional development were important for their teaching contexts. The intention behind including this was to identify what are the importance of assessment among other professional aspects of their teaching practice. Participants categorised these aspects from: 1 = not important and 5 = very important. These CPD aspects were the following: a) language testing and assessment, b) planning lessons, c) knowledge of the subject, d) materials and resources for language education, e) approaches for teaching, f) understanding language learners and needs, g) technology for language education, h) educational policies, i) classroom management, and j) autonomous learning. These were revised throughout the pilot study and one new theme was included (feature j).

Then, the next part of this section dealt with the participation of CPD activities plus its impact on their teaching practice. For item 30 (double-multiple choice item), participants had to select (yes or no) whether they have attended as many of CPD activities listed. These were a) course and workshops, b) conferences, c) qualification programmes, d) online courses and

webinars, e) reading books, f) reading journals and articles, g) mentoring courses, and h) peer observations. The selection of these was based on Richards and Farrell's (2005) taxonomy of professional development activities. These CPD activities were also measured based on the level of impact towards participants teaching practices (being 1= no impact at all to 5= significant impact). Item 31 included some national and international organisations in which participants required to select two responses; the participation or attendance as a nominal measure (e.g., 1=Yes and 2=No) and the impact towards participants' teaching performance, similarly as in item 30. These responses also included one as N/A for those participants who have not attended any CPD, as well as, for the impact of certain institutions in which participants have not been able to attend. A total of 8 institutions were listed: 1) the Mexican association for English language teachers (MEXTESOL), 2) Red Nova, 3) The Mexican Ministry of Education, 4) British Council Mexico, 5) Cambridge University Press, 6) Oxford University Press, 7) Private consultants, and 8) RECALE.

The last part of this section included items (32-34) in relation to participants' preferable learning habits in order to develop the skeleton of the workshop based on how Teachers of Tamaulipas prefer to learn. A total of 4 items were in relation to learning style, activities to enhance learning, length and time for the workshop they would spend. One more item (not numbered) was included to ask participants whether they would consider attending this course for free by providing their emails to be contacted by the researcher at later date. The following section presents the main purpose of the instrument to collect and examined participants' assessment skills and practices.

### **Section 5. Teachers' Assessment Knowledge, Technical Skills and Practices**

A total of 53 double –multiple choice items were included in this section. The structure of this section was based on the three components of the Language Assessment Literacy Standards for language classrooms: Knowledge of assessment (e.g., applied linguistics, theory and concepts and language assessment contexts), technical considerations in language testing and assessment (e.g., testing principles, assessment purposes, assessment process, educational measurement), and school-based assessments ( language assessment, assessment for learning, communication of results, assessment as educational support). This framework was adapted from Zhang and Burrystock (1997), Xu and Brown's (2016) and Giraldo (2018) studies and the items can also be found in Tables 3, 6 and 11. Participants were able to select two responses for each of the items (35-88). The first aspect that participants needed to answer (how skilled you are in...?) and their practices (how often do you do...?). These items were based on how teachers perceive themselves in each of the research framework components. Table 18 presents a table with the research framework of

Teachers Assessment Literacy Development and the sub-themes with the number of items presented in the instrument.

**Table 18 Teachers' Assessment Skills and Practices**

Research framework area	Language Assessment Literacy Standards for Language Classroom: <u>Skills and Practices</u>	No. items	Item number in TALDI
Knowledge of Assessment	Applied linguistics	4	35-38
	Theory and concepts	4	39-42
	Language assessment context	4	43-46
Technical Aspects of Assessment	Testing principles	5	47-51
	Assessment purposes	5	52-56
	Assessment processes	5	57-61
	Educational measurement	5	62-66
School-based Assessments	Language assessment	7	67-73
	Assessment for learning	5	74-78
	Communication of results	5	79-83
	Assessment as educational support	5	84-88

For its final administration and followed the modifications from the recommendations of the Pilot Study, this reliability index (Cronbach's Alpha's) increased up to 0.84, making it highly reliable.

### 3.5.2 Semi-structured Interviews

This method of interviews was suitable for this research study since this investigation aimed at identifying and exploring participants' backgrounds in relation to language assessment literacy and their educational contexts. From Semi-structured interviews is possible to obtain data in which researchers could elicit responses that can be more objective towards the purpose of the research study. Researchers use this type of interview to obtain concise responses by amending the nature of the interviews based on the research aims (Kruger & Casey, 2000).

A large number of research studies in the field of language assessment literacy have used semi-structured interviews in order to collect specific data for participants' assessment beliefs, perceptions and experiences, but especially, to analyse the field of language assessment literacy foundations (Darling-Hammond, 2000; Sikka, Nath & Cohen, 2007; Song & Koh, 2010; Muñoz, Palacio & Escobar, 2011; Jannati, 2015; Seafurrohman & Balinas, 2016; Sheehan & Munro, 2017; Díaz, 2017; González, 2017). As an illustration, Fulcher (2012) and Jannati (2015) developed several interview guidelines to identify, explore and categorise participants' perceptions and experiences in terms of language testing and classroom-based assessment literacies. A disadvantage of conducting research using semi-structured interviews is that the outcome could be manipulated by the researcher in terms of prompting or asking response. Consequently, I have

developed the thematic themes based on TALDI's responses to competent and compare among EFL teachers' perceptions with their educational sectors.

The purpose of including interviews in this research study relied on the factor to further explore TALDI's findings (Phase 1: Survey). It was also included several questions related to the use of language tests and the frequency in which teachers administer these tools to assess students' knowledge. In addition, a section related to problems or malpractices of the use of tests to elicit and identify participants' perceptions among the three different sectors: the private, the public and language institutes. The participants, language teachers and language programme managers were asked about their opinions, practices and experiences about their perceptions towards their language assessment skills and practices, assessment training, and critical issues they have faced when assessment occurs. A total of seven themes were firstly developed and to expand the responses collected from TALDI: a) training in language testing and assessment, b) language tests: importance and frequencies, c) assessment for language skills and linguistic components (grammar, vocabulary and pronunciation), d) knowledge of assessment and technical aspects of assessments, e) school-based assessments, f) teachers' experiences in assessments, and g) issues in assessment within their educational contexts.

Regarding the development of the interview questions, the wording of the interrogative statements did not contain technical lexicon with the intention to facilitate participants to have an appropriate response and avoid any possible bias. An interview guidelines booklet (see Appendix E) has been developed to conduct the interviews to collect participants' views about assessments. This semi-structured interview followed the deductive approach to gather data with the intention of analysing from the general perspective of current trends of assessment of particular cases such as participants' experiences and trajectories (Richards, Roos & Seedhouse, 2011).

Besides this, I had the interview booklet in both languages, Spanish and English. Some interviews were conducted in Spanish for the purpose of teachers expressing their opinions clearer. The methods for conducting the interviews were through face-to-face. Each interview lasts around 15-20 minutes and all interviews were audio recorded with the participants' consent from. Some of the interviews from the pilot study were considered for the final study due to its impact and they provided a fundamental source in understanding the context in depth.

The exploratory research consisted in analysing and examining data from the survey TALDI (Phase 1) and the semi-structured interviews (Phase 2) to conduct a research triangulation to scrutinise the status quo of EFL teachers and language programme managers in relation to their language assessment literacy and educational sectors. The next section describes the instruments for the experimental research: the workshop (Phase 3) and the focus group interview (Phase 4).

### **3.5.3 Instruction of the Workshop in Language Testing and Assessment**

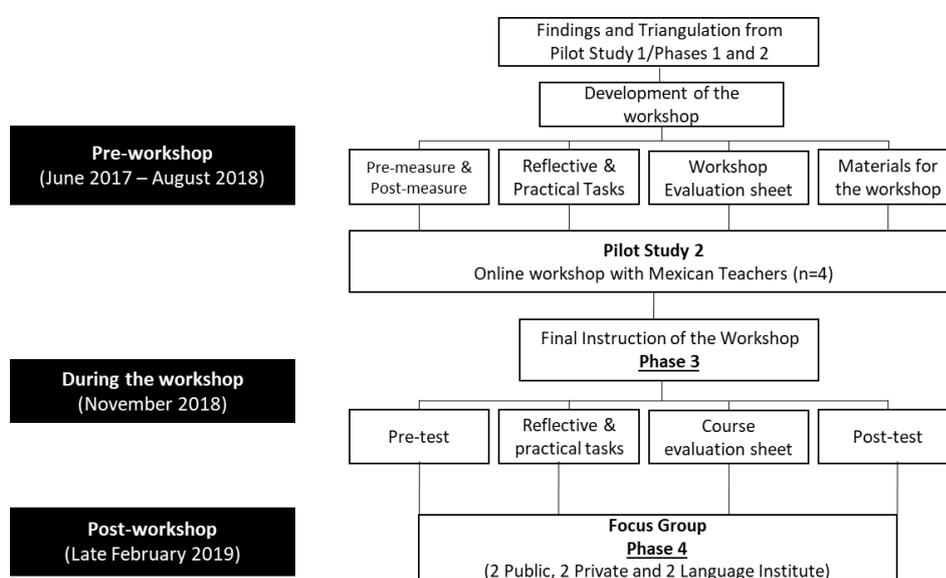
In the field of language assessment literacy, an intervention refers to engage a specific group of participants that were previously selected or randomised based on certain needs that were, to some extent, related to increasing and documenting teachers' assessment practices and skills (Newfields, 2006; Mertler, 2009; Thomas, 2017). A significant number of studies have conducted these interventions to explore teachers' language assessment literacy to discover whether these interventions increase participants' awareness for seeking further assistance to empower teachers' professional development. I decided to develop a workshop for EFL teachers in Tamaulipas to explore their skills and practices, but also to identify participants' experiences and how they treat assessment in their language classroom among different educational sectors. In addition to the data collected from the exploratory research could benefit to notice any discrepancies among the participants' educational sectors.

Language assessment interventions or what language teachers tend to name as a professional development courses or workshops could bring a large benefit to participants, not only in understanding the core of language assessment, but also to apply new (what participants have learnt) knowledge, technical considerations and empower teachers' assessment practices in the classroom. As a result of this, I have included an intervention in this study to investigate how a group of participants from different educational contexts develop their skills in this field, but at the same time to identify their interests towards their continuing professional development. The term intervention has not been included for this research, in fact, it has been replaced as a workshop.

Following the research aims of the experimental research, this workshop aimed at identifying how the participants develop their professional development and the classroom assessment skills and practices and it was analysed in four fundamental aspects: 1) track and measure participants' knowledge in assessment before and after the workshop, analysing the quality of the workshop, 2) gather their perceptions towards the workshop in relation to their current needs and how these had a certain impact on their assessment skills and practices for their language classroom, 3) track participants pathways of development by analysing reflective and practical tasks, and 4) explore the consequential validity of the course towards improving participants' educational sectors.

Therefore, a workshop was needed to be developed based on the current needs that were identified based on the exploratory research from the sample of participants in Phase 1 and Phase 2 of the data collection. Hence, this research study employed a workshop, development of materials, instruction and evaluation stages are presented below.

The following diagram (see Figure 6) highlights the pre-workshop (development), during the workshop (instruction), and post-workshop (evaluation). In the pre-workshop, it was included the interpretation of the data collected from Phase 1 (Survey: TALDI) and Phase 2 (Interviews). This research triangulation delivered the features of what teachers needed for a contextualised workshop to empower Tamaulipas EFL teachers' language assessment literacy. Teachers' needs were collected, and an outline was obtained with different aspects included such as the length of the course, medium of instruction and materials and contents. The next phase was to conduct a pilot study with a small group of Mexican teachers of English (n=4). Based on these two sources of information, the development of the final contents, practical tasks and reflective tasks were formed.



**Figure 6 Development of the Workshop.**

I also included a pre-measure and a post-measure of the course contents. Once all these elements were created and revised, the instruction of the workshop occurred. During the workshop participants took the pre-/post- measure<sup>5</sup> and the reflective and practical tasks. Towards the end of the instruction of the workshop, participants took the post-measure and completed a workshop evaluation sheet. For the post-workshop, a focus group interview was conducted via skype in early February 2019.

### 3.5.3.1 Pre-workshop and Pilot Study 2

With intention of developing the materials, selecting contents and methods for instruction, I was required to analyse the preliminary findings from the final administration of TALDI. According to these findings, 83.6% of the sample (n=263), suggested they would be interested in taking an assessment workshop. This showed an increase from the pilot study 1 with +5.7%. However, 16.4 % indicated they would not be interested in participating in the workshop. The

## Chapter 3

most common reasons among the participants, for not being interested, were the lack of time and interest. One of the sections that TALDI aimed at analysing and gathering information about the participants' preferable methods for learning and their current assessment needs. The starting point was in relation to the learning contents, then methods of instruction, the number of hours participants were required to take part of this workshop and finally participants' suitable professional development activity or learning style, they considered would have a positive impact towards their learning. Table 19 shows the findings for the selection of contents based on what participants found to be interested in for the workshop, this table includes the several themes related to language testing and assessment, the average is reported from a Likert-scale from 1 to 5 with; 1 as not essential and 5 as very essential. The standard deviation was also reported to identify the distribution of the sample. From the list, results showed that most of the participants (N=315) were interested in different topics since the range of average is reported from 2.55 to 4.48. The list of themes has been organised based on participants' priorities and standard deviations were also reported.

Due to the complexity of developing a workshop with all these assessment themes and to enhance and maximise language assessment literacy among the participants, I decided to cluster them into eight different groups based on theoretical and practical aspects. This categorisation also assisted to highlight most of the sessions from the workshop. This categorisation consisted of a) *Assessment for language* (1, 2,3 and 4), b) *Item, task and test development* (8, 12, 15, 18), c) *Purpose of assessment* (6, 11 and 20), d) *Testing principles* (16, 24, 25, 26 and 28), e) *Technical Aspects in Assessment* (9, 14, 17, 21, 27, 30 and 32), f) *Assessment as an educational support* (5, 10 and 22), g) *Specific assessment-related contents* (7 and 31), and h) *Alternative assessments* (13, 19, 23 and 29).

A total of eight sessions were selected based on the final findings. These sessions were divided into general topics: 1) development of language tests, 2) usefulness of tests, 3) purpose of assessment, 4) assessing reflective skills, 5) assessing productive skills, 6) test effectiveness (item analysis), 7) assessment tools, and 8) assessment as an educational support After a revision of the contents, the next stage was to decide the method of instruction and materials to be used. Participants (n=315) showed more interest in the blended learning method (49.3%), whereas the face-to-face method (34.2%) and e-learning (16.4%) obtained fewer responses. As a result, the method of blended learning was the one with more responses and much preferable to gather participants. In relation to the materials in which participants would be able to learn the best, they mentioned that portfolios, presentations, group activities, practical tasks, quizzes and online books were the most appropriate methods to benefit their learning style.

**Table 19 Findings for the Selection of Contents for the Workshop (N=315)**

No.	Language Testing and Assessment Topics related	M	SD
1	Assess productive skills (writing and speaking)	4.48	0.77
2	Assess receptive skills (reading and listening)	4.47	0.75
3	Assess integrated skills	4.45	0.75
4	Assess linguistic features (vocabulary, grammar and pronunciation)	4.42	0.78
5	Interpreting test-takers' scores and performances	4.32	0.76
6	Suitable methods for assessing language	4.30	0.72
7	Assess students with special learning needs	4.26	0.93
8	Development of language tests	4.25	0.78
9	Understanding the Common European Framework bands	4.19	0.92
10	Interpreting test-takers' scores to amend teaching practice	4.16	0.97
11	Types and purposes of tests	4.12	0.82
12	Ethical considerations in testing	4.12	0.94
13	Reporting students' performances (feedback)	4.12	0.99
14	Using holistic scales for productive skills	4.10	0.78
15	Types and purposes of items	4.05	0.83
16	Understanding practicality	4.05	0.88
17	Effectiveness of language tests (item analysis)	4.03	0.90
18	Types and purposes of tasks	4.03	0.82
19	Assess students through self-assessment	4.03	0.73
20	Identify and select topics for testing	4.01	0.96
21	Using analytic scales for productive skills	4.01	0.92
22	Administration of language tests	3.99	0.84
23	Assess students through portfolios	3.97	1.05
24	Understanding validity	3.97	0.97
25	Understanding reliability	3.97	0.88
26	Understanding wash back/ impact	3.96	0.87
27	Writing test specifications	3.95	0.91
28	Understanding authenticity	3.92	0.94
29	Assess students through peer assessment	3.90	0.88
30	Language tests and statistical approaches	3.90	0.84
31	Preparing learners for standardised tests (FCE, etc.)	3.90	1.00
32	History of language testing	2.55	1.08

The number of hours played an important role in the distribution of the contents and how participants would interact during the workshop. Based on the TALDI, participants indicated they would be able to participate in the following range of hours, from 0-10 hours (34.2%), 11-19 (15.1%), 20-29 (26%), 30-39 (2.7%), and above 40 hours of training (15.1%). These results showed that the workshop would be around 10 to 20 hours at least for attending and around 10 hours of independent work. The next stage of the development of the background of the course to develop the materials such as power-point presentations, practical tasks, and reflective tasks to be tested in an upcoming pilot study with a smaller group of EFL teachers in Tamaulipas.

Once the 8 sessions were categorised, the next step was to finalise what EFL teachers in Tamaulipas were required and highlighted differences among participants' educational sectors.

### Chapter 3

A smaller pilot study was conducted to examine the materials and the contents of the workshop. Participants were contacted through emails that they provided during the pilot and final administration of the TALDI. A total of 43 emails were sent and only 19 replied to inform they were interested. An online platform was used to include the contents of the eight sessions and materials were developed in relation to their topics. There were three sessions with the participants via Skype. This pilot study lasted two months in total due to a lack of participation in completing the activities and delivering the practical and reflective tasks. There were only four participants who fully completed the materials and were involved in the three Skype sessions.

At the end of this pilot study, I collected participants' perceptions based on their comments. Some of the aspects they suggested that require improvements were related to the methods of the instruction. They indicated they would prefer to have a face-to-face workshop because of the lack of time and confidence in technological knowledge towards the use of the platform that was used. Another suggestion was that the contents of the course were not that contextualised; in other words, some of the contents were not that applicable to the sample of participants (Tamaulipas EFL teachers) from the pilot study. As a result, these findings from the pilot study flagged up essential considerations to be discussed in relation to the Mexican educational context. Besides this, one participant suggested that it would be beneficial to identify their progress somehow with a simple and short quiz at the beginning and/ or at the end that could be administered to let them know their general progress.

Based on the comments received and the lack of time available from the participants, the methods for the delivery of content changed into face-to-face for the final instruction of the workshop. This action suggested that the final workshop should take place in the Mexican State of Tamaulipas and the blended learning method should be discarded. With regards to the lack of contextualised concepts, a new session was developed, "*Problematic situations in Assessment in Tamaulipas within language assessment contexts*". This would be the most challenging aspect since it includes what EFL teachers currently require, but with the data from the semi-structured interviews and TALDI would be possible to obtain vital features to develop this session.

In addition, a measure was developed with the relevant theoretical and practical considerations for the course. The purpose of this measure was to identify what the participants knew before and after the course. This measure was administered at the beginning and the end, this instrument was the same in both administrations, but the order of the items changed. This could be an essential aspect that was not considered before the pilot study, but it could bring a significant number of advantages at the moment of measuring and tracking participants' progress and interests throughout the workshop.

Overall, this pilot study was essential to identify the different methods in which participants considered certain aspects essential and to increase the level of contextualisation (what was important for them in Tamaulipas). It is worth noting that the contextualised workshop refers to critical aspects of the use of assessment in different institutions in Tamaulipas. The course had been developed to discuss, identify and provide solutions from participants' current issues among the public, private and language institute sectors. The following section presents the instruction of the workshop and several aspects such as the syllabus, the measure, the reflective and practical tasks and the course evaluation sheet.

### **3.5.3.2 During the Workshop**

In the first session, participants received a folder with all the materials including the syllabus, the power-point presentations (10 in total) printed and space to write their notes, the information about practical and the reflective tasks. I gave instructions about the workshop purposes and the methods in which participants should complete each of the tasks. All participants signed the consent form. This section included information about the instruction of the workshop in relation to the syllabus, measures, the tasks and the workshop evaluation sheet. The course lasted slightly more 20 hours of instruction and 10 hours of independent practice. The following subsection discusses the syllabus, pre-/post- measures, reflective and practical tasks, and the course evaluation sheet.

### **3.5.3.3 Syllabus**

The workshop was divided into 10 sessions. Each session lasted two hours approximately in which participants were able to discuss and gain knowledge, skills and practices on different concepts and themes related to language testing and assessment. The ten sessions were 1) introduction to the workshop, 2) testing, assessment and evaluation in language education, 3) language testing, 4) language assessment standards, 5) language assessment in receptive skills, 6) evaluation of items and test, 7) language assessment in productive skills, 8) communication of assessments and error correction, 9) alternative assessments, and 10) language testing in Mexico and Tamaulipas. The full version of the syllabus from this contextualised workshop can be found in Appendix H. In addition, Participants took the pre-measure, tool that was developed based on what participants from the pilot study 2 recommended. The following section describes the contents of the tool and how these contents and items were related to the sessions of the workshop.

#### **3.5.3.3.1 (Pre- and Post-) Measures**

The initial measure was administered in the first session. This 40 multiple-choice item tool was developed with the intention to collect participants' knowledge at the beginning of the workshop. This instrument was also administered at the end of the course to illustrate and track

the amount of (increased) knowledge that participants had gained throughout the workshop sessions. This instrument included a total of 9 topics related to the field of language testing. Table 20 presents the contents of this instrument, the number of items and how these items were related to the workshop sessions. This tool has a limitation in which its validity and reliability were not statistically analysed, but it served the purpose of guiding participants and me to highlight which areas they made more progress and were more interested. Besides this, this instrument was useful to highlight different areas and to link to the standards for language assessment literacy in the classroom.

**Table 20 Contents and Items of the Measure**

Language Assessment topic	# items	Workshop session
Important Concepts in Assessment	3	2
Testing principles	5	3 and 4
Development of test: Types of items and tasks	11	3, 4, 5 and 7
Use of marking criteria	3	7
Communication of results and correction	3	8
Alternative assessments	5	9
Item analysis	4	6
Assessment as educational tool	3	6 and 9
Mexican Educational Context	3	10
Total	40	

The format of this instrument was similar to the Teaching Knowledge Test (TKT) which participants were familiar to. Items are categorised per themes and include multiple-choice items that participants were able to match based on a statement that participants had to choose the appropriate answer. This was piloted with 5 teachers to avoid any bias and made the statements logical and coherent. The full version of this tool can be found in appendix H with its answer key.

### **3.5.3.3.2 Reflective Tasks**

The purpose of the reflective tasks was to identify how participants reacted to some of the contents from the workshop and these were useful to collect their perceptions towards the development of knowledge among participants sectors. Xu and Brown (2016) indicated that reflection was an essential aspect for teachers to identify and therefore to develop new skills and practices. A total of 3 reflective tasks were administered throughout the 10-session workshop. Participants had to write each of these 3 reflections with no more than 150 words either in English or Spanish, based on their language of confidence. The table below (see Table 21) presents the reflective tasks and how these ones are related to the workshop contents.

**Table 21 Reflective Tasks and Workshop Sessions**

Task	Reflection purpose and questions	Session related
1	<u><i>Important considerations in language assessment for the classroom</i></u> 1. What is assessment to you and what is its importance in the language classroom? 2. How do you conduct assessments in your classroom? 3. What do you think the best methods for assessments are?	2
2	<u><i>Development of tests and efficiency</i></u> 1. What elements do you consider essential for the development of language tests or testing instruments? 2. What are your perceptions towards the use of quantitative research methods to maximise language tests? 3. What sort of information can you collect if you conduct item analysis?	3, 4, 5 and 6
3	<u><i>Problems and solutions for Language assessment in the Tamaulipas classrooms of English</i></u> 1. What do you consider problematic at the moment of using assessments in your classroom? 2. Can you provide one or more solutions to the problem(s) you have identified?	10

These reflections were written in paper after session 2, 6 and at the end of the workshop. Each reflection dealt with one specific topic: 1) important considerations in language assessment for the classroom, 2) development of language tests and efficiency, and 3) problems and solutions for language assessment in Tamaulipas classrooms of English. Each of these tasks included open-ended questions so, participants could write their opinions and perceptions.

### 3.5.3.3.3 Practical Tasks

There were two practical tasks conducted during the workshop. Table 22 shows these two tasks and the activities in which participants had to achieve. The first task was a group activity in which participants had to develop a test and conduct an item analysis to discover the efficiency of the items. Allowing participants among different sectors could be an opportunity to identify different aspects based on their workplace. The second task was an individual analysis of one aspect of language assessment in the classroom such as communication of results or alternative assessments.

**Table 22 Practical Tasks and Workshop Sessions**

Task	Objectives	Session related
1	A group practical task, participants were able to: a) Develop a task specification of one section of the test related to reading, listening or grammar and vocabulary, b) Develop adequate items for the section c) Administer the test with a small sample of test takers d) Input the data in Excel and conduct item facility, item discrimination and distractor analysis e) Interpret the results	2-6
2	An individual task, participants were able to: a) Evaluate a tool or an aspect which participants consider essential for assessment in the classroom	7-9

## Chapter 3

The first practical task was a group activity in which participants had to develop a test from scratch, creating a task specification detailing all aspects of the test, developing and choosing appropriate items, conducted the tests in their classes and then analyse the efficiency of the items through item analysis. Participants worked in different groups to develop 2 language tests. One test was aimed at level A2 and the other B1. Each test included several items divided into reading and listening comprehension, as well as grammar and vocabulary. The number of items per test section was decided by the participants.

The participants (n=10) were divided into groups of 3 to develop each section of the test. For the test of A2 (CEFRL), 2 participants developed the grammar and vocabulary section, 2 participants for A2 Reading and 1 participant for listening: similar process for the test B1 (CEFRL). For the purpose of this task, they had to develop a task specification sheet in which they input the essential aspects of the tests and the purpose of such testing instruments. Participants put into practice different aspects of test development and item analysis. Once they complete the tasks specification and had developed the tests, these were administered in their institutions with a small sample of their students.

Since participants had just one day to administer the test and input the responses (in a spreadsheet I developed to facilitate this process) for item analysis, the number of tests administered was 23 for A2 and 14 for B1. Participants input the data into a Microsoft Excel spreadsheet in which they used the knowledge learnt from session 6. They conducted an item analysis and identified which items were useful for the level and purpose, as well as participants were able to evaluate the functionality and practicality of the items and finally discarded the ones not useful for the test.

The second practical task was to evaluate one aspect discussed in the workshop that was related to language testing and assessment (allowing participants to show their interest and elicit their needs in different educational sectors), the aspects chosen by the participants were among communication of assessments, use of criteria for productive skills and alternative assessments (e.g., language portfolios). Participants were free to select at least one aspect and analyse it based on the positive and negative impact towards students' performances in their educational sectors. Participants discussed how this aspect would benefit their assessment skills and practices in the language classroom as an individual presentation in front of other participants and the researcher.

### **3.5.3.3.4 Workshop Evaluation Sheet**

A workshop evaluation sheet (WES) was developed to obtain participants' experiences throughout the two-week workshop (this can be found in Appendix J). The first part (A) focuses on

collecting evidence of five elements: a) contents of the intervention, b) instructor performance, c) evaluation/ usefulness of reflective and practical tasks, d) materials used for the intervention, and e) overall perspective. Each of the sections has several items (statements) that participants could self-evaluate based on their performance and involvement during the workshop by using a Likert-scale. The second part (B) was an open-ended question section to obtain more information on how the course should improve to maximise participants' professional development. Overall, during the workshop (Phase 3), there were four types of instruments which participants completed, these were the measures, practical tasks, reflective tasks and workshop evaluation sheet. The following section provides information on how the post-workshop took place via a focus group interview (Phase 4).

#### **3.5.4 Focus Group Interview**

A focus group interview is a useful qualitative source of data to fill gaps among a group of participants who have interests for a particular topic (Cohen, Manion and Morrison, 2011). This type of interview could assist researchers to observe and highlight several aspects that need further attention and examine in a deeper way by sharing ideas and reflections. A focus group has a significant number of advantages to seek further evidence of an experiment. One of these advantages is that a focus group could provide current information about a specific aspect being analysed. In addition, focus groups are friendly user since participants can interact to each other in an open discussion. However, one disadvantage could be that data might not be useful if the focus group is not well-oriented to the purpose of the research (Hernandez, Fernandez & Baptista, 2003).

The post-workshop stage consisted in conducting a Focus Group Interview (Phase 4) with a small sample of participants from the workshop. This aimed at examining the progress of their assessment skills and practices, as well as in which areas of assessments participants developed interested could potentially provide more discrepancies among participants who took the workshop towards their assessment literacy and their educational sectors. A focus group is an opportunity to express different opinions among participants who were part during the workshop. Therefore, having participants from the three educational sectors could bring certain differences that I was not aware of at the moment of delivering the workshop. In fact, this could also bring new aspects for the exploratory research conducted in Phase 1 and 2.

Consequential validity in this research explores the impact, either positive or negative, on how participants perceived during the workshop towards their educational sectors. Also, to measure the consequential validity of the course in relation to how participants could have applied the assessment skills and practices they had previously learnt and discussed during the

workshop. Participants were asked about how they had used different strategies, skills and practices, they acquired in the intervention and to which extent they have been useful. A list of questions was developed with the intention of gathering information about their current practices and whether there had been, if any, changes in the use and experiences of language assessments in their educational settings. The four data collection phases, the survey (Phase 1), semi-structured interviews (phase 2), materials and instruments used in the workshop (Phase 3) and the focus group (phase 4) were analysed by mixed-method and concurrent-mixed methods approach in which the statistical tests were conducted and thematic analysis were emerged from the semi-structured interviews and focus group.

### **3.6 Data Analysis Methods**

The section presents the data analysis methods for each of the four phases: TALDI, semi-structured interviews, workshop materials (measures, reflective, practical tasks and course evaluation sheet) and the focus group interview. Also, it will introduce how the data was input, categorised, analysed and interpreted from the exploratory and experimental research methodologies. The Phase 1 (Survey) and 2 (Interviews) data were collected to answer research questions related to the status quo of EFL teachers and language programme managers in Tamaulipas (research question 1 and 2) and professional development trajectories (research question 3). The Phase 3 (Workshop) and 4 (Focus Group) data were analysed based on the participants who demonstrated that had the most and least progress and interest (research question 4) and impact of the workshop on the participants (research question 5).

#### **3.6.1 Phase 1 (Survey: TALDI)**

The survey is one of the features for the mixed-method analysis for the exploratory research. The quantitative data was initially input and categorised in Microsoft Excel (2013) spreadsheet with the intention to clear the data, then it was exported onto the Statistical Package for Social Sciences (SPSS v.25). Descriptive statistics were applied to identify differences in relation to participant's backgrounds including educational sectors, interests in testing and assessment, also their professional development areas that they considered important to their teaching practice. Standard deviation and percentages were reported also based among four sections of the Survey (TALDI): 1 (participants' personal information), 3 (Second Language Testing and Assessment), 4 (Participants' Continuing Professional Development) and 5 (Participants' assessment skills and practices) of TALDI.

Since one of the main research aims is to discover discrepancies, if any, among the three different cohorts of participants (the public, the private and language institutes sectors)

inferential statistics were also conducted to compare these three different educational sectors. These statistical tests were applied to analyse the participants' assessment skills and practices (TALDI: Section 5) from the eleven standards of language assessment literacy (see Tables 3, 6 and 11). There were three components from the theoretical framework, **Knowledge of Assessment** that included the standards of *Applied Linguistics* (AL), *Learning and Assessment Theories* (LAT) and *Institutional Assessment Policies* (IAP). The component for **Technical Consideration in Assessment** included the standards of *Testing Principles* (TP), *Assessment Purposes* (APu), *Assessment Processes* (APr) and *Educational Measurement* (EM). The last component relates to the **School-based Assessments**, which includes the standards of *Language Assessment* (LA), *Assessment for Learning* (AfL), *Communication of Assessments and Results* (CAR), *Assessment as an Educational Support* (AES).

The first statistical test was Pearson Correlations ( $r$ ), a parametric test which measures the relationship of two variables. These results are usually reported based on positive ( $r=1.00$ ) or negative ( $r=-1.00$ ) relationships between two variables. Pearson correlations were calculated to identify relationships among the participants' assessment skills and practices among the standards based on participants' language assessment literacy. They were calculated also to identify any relationships among participants' educational sectors concerning independent variables such as participants' assessment experiences, academic and professional trajectories, interest in language testing and assessment, and previous training experiences.

Analysis of Variance (ANOVA) was applied which aimed at comparing two or more groups (educational sectors) among the eleven different aspects of the language assessment literacy framework: participants' assessment skills and practices. The means of skills and practices per standard were compared among the three groups to find statistical significance among the educational contexts. After conducting ANOVA, to maximise and to be sure about differences among the three sectors, Tukey analysis was also conducted. Tukey test or honestly significance difference (HSD) was also administered to the ANOVA tables in SPSS. This test is vital for identifying differences among the groups from the variables that have been studied (standards of Language Assessment Literacy).

In addition, this survey included a qualitative section (TALDI: Section 2) that contained within a total of seven open-ended questions. Participants' entries were analysed based on their frequencies of emerging themes on participants' assessment experiences in their educational sectors. This was possible due to the transcription and analysis of the repetitive answers from the participants (frequencies). A total of six themes emerged from the data collected: a) test development and responsibilities, b) assessment features, c) alternative assessments methods

and efficiency (this was merged due to similarities in participants' responses), d) ethical considerations in assessment, e) teachers' previous and current assessment experiences and f) critical issues in school-based assessments. The findings from this survey can be found in Chapter 4.

### 3.6.2 Phase 2 (Interviews)

This is the second source of data for the mixed-method approach for the exploratory research. The data from the TALDI and the semi-structured interviews was triangulated to seek further discrepancies. The analysis of the semi-structured interviews was based on a thematic analysis with the intention of identifying participants' professional development, their academic background, the institutional environment and educational sectors. The interview data were grouped based on educational contexts and fully transcribed. Translations were made by the researcher for those interviews that were conducted with the participants' first language (Spanish). Firstly, EFL teachers' interviews were transcribed and analysed. Then, the language programme managers' perceptions were transcribed, analysed and compared to teachers' perceptions and to support and scrutinise their educational sectors.

The use of NVIVO v.11 was vital with the intention to explore and categorise the emerged themes with the purpose of comparing the outcomes with the findings from the TALDI. Seven themes were emerged based on the findings. Table 23 shows how the 7 themes from the semi-structured interviews were examined and contrasted with the findings from the TALDI. A methodological triangulation was conducted from Phase 1 and 2 to answer research questions 1, 2 and 3 that were related to the status quo of language assessment literacy among EFL teachers in Tamaulipas and their academic and professional trajectories.

**Table 23 Triangulation of the Mixed-method Approaches**

Themes for the semi-structured Interviews (Phase 2)	TALDI Sections (Phase 1)
1) Training in language testing and assessment	3 and 4
2) Language tests: importance and frequencies	2
3) Assessment for language skills and linguistic components	
4) Language Assessment knowledge and technical skills	1 and 5
5) Language assessment practices in the classroom	
6) Language assessment teachers' experiences	
7) Problems in their teaching contexts	2

The analysis was based on different codes from NVivo that allow me to separate and differ from different aspects of the interviews. Three other themes were emerged based on the support from the language programme managers. There were more related to the development of language tests since some of the teachers were not responsible for developing these learning

tools. The other theme was in relation to the code of practice and the ethical considerations towards avoiding bias during the test. Lastly, one more theme was discovered during marking and delivering results to students. The findings from the semi-structured interviews are presented in Chapter 5. In addition, a summary of the findings from the exploratory research can be found at the end of Chapter 5

### **3.6.3 Phase 3 (Workshop)**

The workshop was vital in order to conduct the experimental research. The analysis of the workshop materials (e.g., pre- and post- measure, reflective tasks, practical tasks and workshop evaluation sheet) were through a concurrent mix-method analysis in which aspects such as educational sectors, level of confidence and issues related to their teaching contexts were considered essential. The scores from the measure were input in Microsoft Excel (2013) spreadsheet to identify the differences in their scores. Descriptive statistics were reported. Data from the workshop evaluation sheet was analysed by observing changes in the averages of means from the different aspects of the workshop such as contents and objectives, learning resources, activities, discussions and tutor performance (the researcher). These two sources were analysed quantitatively. The use of SPSS was discarded due to the small sample of participants to conduct inferential statistics; thus, differences were reported based on means.

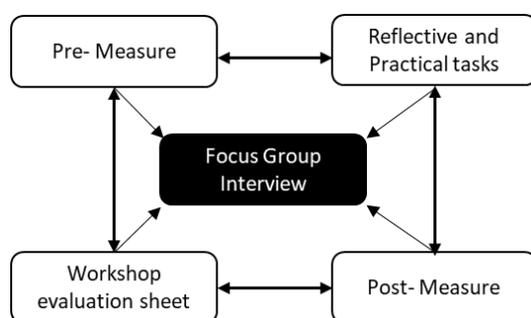
The practical and reflective tasks were analysed by identifying similarities in themes that could also corroborate that participants were learning and expanding their assessment skills and practices. These were scrutinised based on participants' opinions and perceptions towards the purpose of the tasks. The findings from these two types of tasks were essential to investigate how participants from different educational sectors worked and performed throughout the course. These findings were reported based on their educational sectors.

### **3.6.4 Phase 4 (Focus Group)**

The last part of this research project was to conduct a focus group to explore the impact that the contextualised workshop might have on participants. Some participants communicated their ideas in both languages, English and Spanish; as a result, the focus group interview was transcribed and translated into English. The use of NVivo was vital to highlight the different emerging themes and therefore, develop a coherent method for its analysis and interpretation.

This analysis was reported based on four dimensions. The first one was to investigate whether the course had an impact on the students based on how teachers had performed after attending the workshop. The second was similar to the first but considering how teachers perceived themselves based on their previous experiences and what they had learnt after the

workshop. The third dealt with the institutional perspective. In addition, the purpose of this focus group was also to compare the findings from the worksheet evaluation sheet. These findings were based on the concurrent mixed-methods, to complement and support participants' statements. Although no language programme managers participated in the focus group nor during the workshop, participants were able to identify tensions between these two stakeholders at the institution where they work. Finally, the fourth dimension dealt with the impact on professional, society and the political aspects from the federal government and the current educational reforms in Tamaulipas, as well as in Mexico. Figure 7 illustrates how phase 3 and 4 were analysed and findings were reported. Chapter 6 presents the findings with the participants' data collection as reflections (REF), workshop evaluation sheet (WES) and focus group interviews (FG).



**Figure 7 Data Analysis of Phase 3 and 4**

### 3.7 Summary of the Chapter

This chapter presented an overview of the research project with the methods applied for data collection and how they were analysed. This investigation followed two foci: exploratory and experimental research. Ethical aspects and approval were considered during this research project by storing participants' personal information. The exploratory investigation was conducted with EFL teachers and language programme managers with the use of the Survey (TALDI) and semi-structured interviews. A total of 315 participants took part in the first phase of the data collection. The TALDI findings were triangulated with the semi-structured interviews (n=15). This mixed-method approach was useful for answering research questions 1 to 3.

The experimental research was conducted with a group of EFL teachers who attended a workshop related to language testing and assessment and then a focus group. The third and fourth data collection phases were based on the experimental approach in developing the contextualised course for teachers in order to enhance their language assessment literacy. Participants (n=10) from the instruction of the workshop had a total of four main aspects were analysed by concurrent mixed-methods: the pre- and post-measure, the reflective and practical

tasks and the course evaluation sheet. The purpose of analysing these four was to investigate participants' progress and which features participants were less and more interested during the intervention. In relation to the fourth phase, a focus group was conducted with six participants, 3 from each sector who took part in this workshop. The data was triangulated through mixed-methods approaches to measure the impact on the participants from the course. Four aspects were analysed: how the course might have an impact on participants' students, personal trajectories, institutional level and socio-political aspects. A research study overview is presented in the following table (see Table 24) with the research questions, research aims, and data collection procedures and data analysis.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

**Table 24 Research Study Overview**

Research Question	Research Aim	Data sources	Data Collection Phases	Data Analysis
<b>1. What is the <i>status quo</i> of EFL teachers' language assessment literacy in Tamaulipas?</b>	This research question aims to explore and analyse EFL teachers status quo of language assessment literacy	Test relevance and occurrences, test development and responsibilities, assessment of language skills and linguistic components, Self- perceived teachers language assessments skills and practices (TLASP)	<b>Phase 1:</b> (Survey: Teachers' Assessment Literacy Development Inventory)  <b>Phase 2:</b> (Semi-structured interviews)	Descriptive statistics Pearson Correlation Multiple Analysis of Variances (ANOVA) [3 educational sectors] and thematic analysis
<b>2. What are the differences, if any, in EFL teachers' language assessment skills and practices among different educational sectors in Tamaulipas?</b>	To discover any discrepancies on EFL teachers' language assessment literacy among educational sectors (the public, private and language institutes)			
<b>3. What are EFL teachers' and programme managers' backgrounds in their professional development towards language assessment literacy?</b>	To identify and analyse the professional development trajectories of EFL teachers. Also, this research question explores and examines their previous and current assessment experiences, academic and professional trajectories	EFL teachers' assessment experiences, as well as their academic, practicum and professional trajectories.		Thematic Analysis
<b>4. In what ways has the workshop influenced EFL teachers' awareness of language assessment literacy?</b>	This research questions explore how EFL teachers progressed in the workshop	Workshop materials: pre- and post-measures, reflective and practical tasks (REF), workshop evaluation sheet (WES). In addition, the focus group interview (FG) was conducted to further explore participants' progress, as well as the impact of the workshop on EFL teachers.	<b>Phase 3</b> (Workshop)  <b>Phase 4</b> (Focus group interview)	Descriptive statistics for the measures and the workshop evaluation sheet  Thematic analysis for the reflective and practical tasks, as well as for the focus group interview
<b>5. What was the impact of a workshop of language assessment literacy development on EFL Teachers in Tamaulipas educational sectors?</b>	To measure the impact of a contextualised workshop on EFL teachers. Also to develop a checklist where teachers could self-evaluate what they can do in different standards of language assessment literacy development			

## Chapter 4 Survey (TALDI) Findings

This Chapter aims to demonstrate the findings of the quantitative instrument, Teachers' Language Assessment Literacy Development Inventory (TALDI) administered among EFL teachers in Tamaulipas. This section has been divided into six different sub-sections. The first sub-section reports on participants' academic trajectories, the number of students allocated in their classrooms and the level they were currently teaching. Section 4.2 reports of the questions from section 2 of TALDI which were seven open-ended questions related to participants' schools' backgrounds and this demonstrated a total of six themes to explore. Section 4.3 reports on the participants' professional development in the area of language testing and assessment, as well as the interest in different topics related to this field. Section 4.4 reports on participants' general continuing professional development, as well as their preferred methods for learning so as to develop a workshop with their current interests and needs. Section 4.5 provides the findings of the section 5 from TALDI which were analysed based on the statistical test Analysis of Variance (ANOVA). Finally, this Chapter concludes with a summary including the most significant findings from TALDI analysis.

### 4.1 Participants' Educational Backgrounds

In relation to the participants' academic trajectories, data highlighted that from the total sample (n=315), 204 had been awarded with Bachelor's degrees (64.7%), 49 with Diploma or Specialisation degree (15.5%), 61 with Master degrees (19.3%) and one participant with a degree of Doctor in Philosophy (0.05%). For the public sector, there were 76 participants with bachelor's degrees, 7 with diploma courses and 22 with master's degrees. The private sector, 57 with bachelor's degrees, 24 with diploma courses and 27 with master's degrees. For the language institutes, 71 bachelor's degree, 18 with diploma courses, 12 with master's degrees and one with a doctorate degree.

Table 25 presents each of the educational sectors and the number of participants per academic disciplines. Participants' academic profiles were mainly the English Language Teaching (ELT) and Teaching English for Speakers of Other Languages (TESOL) programmes with a total of 114, being the most common academic programmes between the participants from the public and private sectors. On the other hand, Applied Linguistics (n=42) was in most common within the language institutes with 74 participants. Data also reported that the public and private sector tented to hire EFL teachers with other disciplines that were not linked to language education. The public sector captured a total of 12 participants whose academic trajectories were not directly

based on a language teaching background. Participants from the private sector (n=29) also did not have a direct connection between their academic disciplines and the English courses they were currently teaching. A total of 13 participants (n= 1 public and n=12 private) indicated that they had studied other disciplines such as Engineering, International Law, Business and Administration, Divinity and Veterinarian studies.

**Table 25 Participants' Academic Profiles**

	Applied Linguistics	Pedagogy	ELT/ TESOL	Elementary Education	Others in Education	Others	Total
Public Sector	15	15	41	22	11	1	<b>105</b>
Private Sector	17	5	36	21	17	12	<b>108</b>
LI	42	19	37	1	3	0	<b>102</b>
<b>Total</b>	<b>74</b>	<b>39</b>	<b>114</b>	<b>44</b>	<b>31</b>	<b>13</b>	<b>315</b>

This diversity of academic backgrounds might lead to certain issues during conducting teaching and assessment practices in the language classroom due to a lack of academic preparation. It was also important to consider that continuing professional development could potentially enhance teachers to empower their practices inside the language classroom.

The next aspect analysed was participants' number of allocated students per classroom and the number of teaching hours per week. Seventy-six participants (72.4%) from the public sector indicated that the number of students in a language classroom was among 31-40 and 25.7% (n=27) indicated that the number of students taught could reach more than 40 students in a public institution. Only two participants from the same cohort indicated that they were teaching in classrooms with 21 to 30 students. In terms of this sector's teaching hours per week, 24 participants (25.2%) indicated that they worked from 21-29 hours and 29 participants (30.4%) from 30-40 hours. The public sector also reported a short period of teaching hours, 6 to 10 with 14 participants (13.3%), these were mainly from part-time positions.

For the private sector, 71 participants (65.7%) indicated that the number of students in a classroom was among 21-30 and 30 (27.8%) reported that the number of students was higher 11-15. Only seven participants (6.5%) indicated they have around 31-40 students per classroom. In terms of teaching hours per week, 61 participants (56.4%) indicated that they usually spent on average 16-20 hours a week, 25 (23.1%) around 21-29 hours, 14 (12.9%) indicated they were teaching from 11-15 hours.

For the language institute cohort, 75 (74.5%) indicated that they taught 21-30 students in a classroom, 24 participants (23.5%) with 11 to 15 students and three participants more than 30 students. EFL teachers from this cohort, tend to work the shortest time of teaching hours per week. Four participants indicated that they worked from 1 to 5 hours a week. The majority of language institute participants worked from 16-20 hours with 81 participants (79.4%). This feature

was vital to examine the possibility of malpractice of how to conduct classroom assessments since the number of students is higher and teachers can also teach one or more groups with the same number of students.

In relation to participants' teaching contexts based on the CEFRL, data reported that participants from the public sector pointed out that levels A1 and A2 were the most common with 50 (47.3%) and 51 (48.1%), respectively. Level B1 was captured with 3 participants (2.9%) and 1 for B2 (1.7%). For the private sector, the English language level of teaching contexts was reported higher with 49 (45.4%) and 41 (38%), B1 and B2 correspondingly. The other 16.6% was distributed with 13 participants teaching A2 and 5 participants at C1. Participants from the private and language institute showed an equal distribution across the intermediate (B1 and B2) CEFRL levels: 43.3% for B1 and 40.2% for B2.

The findings from TALDI (Section 1) were significant in terms of identifying participants' backgrounds in relation to their educational sectors, participants' academic trajectories and disciplines of studies, as well as their classroom settings (classroom size) and the level of the language courses they were currently teaching. Some of the relevant findings were that some teachers do not have relevant academic backgrounds in terms of the contents and modules they were currently teaching. Multi-disciplinary teachers, those whose academic trajectories do not reflect the contents they are currently teaching, were mainly found in the private sector. A total of 271 showed that they had previously been awarded with academic programmes related to English language teaching. An important finding was the higher number of students in the public sector compared to the other two sectors being examined. Also, intermediate levels were commonly taught for the private and the language institute sector. The following section presents the findings from the participants' responses in relation to their assessment experiences in their educational sectors and their institutional assessment policies.

## **4.2 Participants' Assessment Experiences and Educational Sectors**

The second section of TALDI (*Participants' Assessment Experiences in their Educational Sectors*) had seven open-ended items, these participants' responses were analysed based on a thematic analysis: a) test development and responsibilities, b) assessment features, c) alternative assessments methods and efficiency (this was merged due to similarities in participants' responses), d) ethical considerations in assessment, e) teachers' previous and current assessment experiences and f) critical issues in school-based assessments.

### 4.2.1 Test Development and Responsibilities

There were 268 responses that were captured. The majority of the participants (n=163) commented they were not responsible for the development of language tests. They responded that it was the responsibility of the language programme managers to develop, review and hand the tests into the teachers. A total of 105 teachers indicated that they develop their tests or uses other sources to adapt them into their testing resources. Notably, there were seven participants who claimed that they did not use tests as part of their institutional assessment policies. From this sample, participants (n=33) wrote that they had balanced methods of assessments in which some summative tests were given and that they permitted to develop formative tests. Table 26 highlights the participants' educational sectors and the test development responsibilities

**Table 26 Test Responsibilities among Educational Sectors (n=268)**

Test Development	Public (n=85)	Private (n=95)	Language Institutes (n=88)	Total
Teachers	65 (76.4%)	20 (21.1%)	13 (14.7%)	<b>98 (36.5%)</b>
Language programme managers	16 (18.8%)	72 (75.8%)	75 (85.3%)	<b>163 (60.8%)</b>
No use of tests	4 (4.8%)	3 (3.1%)	0	<b>7 (2.7%)</b>
<b>Total</b>	<b>85 (31.8%)</b>	<b>95 (35.4%)</b>	<b>88 (32.8%)</b>	<b>268</b>

In relation to participants' educational sectors, the public sector (n=85) commented that 65 developed their tests for their classes. From this previous sample, 19 informed that the tests they administered had been developed by the course books they were using in their classrooms, and 48 indicated that some tests were extracted from different sources as books, internet and other teaching resources. However, 16 respondents mentioned that tests were given by the school authorities. There were four from this sector who did not use tests as part of their assessment process. For the private sector, a total of 95 responses were captured: 20 participants indicated that they developed their own language tests and 72 responses reported that tests were developed by their language managers. There were no responses from the private sector in relation to what they take into consideration to develop their tests. Similarly, as the public sector, there were three participants who did not use tests as part of their evaluation. For the language institutes, a total of 88 responses were recorded; 75 participants indicated that the tests were developed by their language programme managers and 13 that they use mock tests, they mentioned that they mainly use standardised tests and resources from books to develop their own language tests. The following section presents the findings exploring what participants currently test in their school-based assessments.

#### 4.2.2 Language Skills in Assessments

This theme obtained a total of 256 responses. Table 27 summarises the data analysis based on the three educational sectors and the language skills (i.e., reading (R), listening (L), speaking (S) and writing (W) and the use of tests for assessing students' grammar (G) and vocabulary (V).

**Table 27 Matrix of Language Skills within Tests (n=256)**

Language Skills and Linguistic features	Public (n=84)	Private (n=94)	Language Institute (n=78)
G,V;R,W,L,S	5 (5.9%)	23 (24.5%)	55 (70.5%)
G,V;R,W,L	12 (14.3%)	0	0
G,V;R, L	23 (27.4%)	0	0
G,V;	44 (52.4%)	0	0
R,W,L,S	0	27 (28.7%)	23 (28.5%)
R,L,S	0	12 (12.7%)	0
R,L	0	32 (34.1%)	0

Participants from the public sector recorded a total of 84 responses. Five participants (5.9%) indicated that they assessed the four language skills and grammar and vocabulary. Twelve participants (14.3%) stated that they did not include speaking in their assessment, 23 (27.4%) excluded writing and speaking, and 44 participants (52.4%) only assess students' grammar and vocabulary. The private sector reported a total of 94 entries in which 23 (24.5%) indicated they assessed the four skills and grammar and vocabulary, 27 (28.7%) only assessed the four language skills, 12 (12.7%) excluded writing and 32 (34.1%) excluded writing and speaking. Participants from the language institutes recorded 78 responses: 55 responses (70.5%) were in relation to the four language skills and grammar and vocabulary and 23 (28.5%) participants stated that they only assess the four language skills.

The discrepancies on how teachers assess different language skills tended to vary among the three educational sectors. Participants from the public sector showed that their students take more tests in relation to grammar and vocabulary. The private sector and language institutes tend to cover most of the skills, but there is a lack of speaking assessments from the private sector. Language institutes tend to develop tests with the four language skills and grammar and vocabulary. The following theme presents the findings of the methods of classroom assessments, those excluding language tests.

#### 4.2.3 Alternative Assessments

For this theme, a total of 237 commentaries were collected and analysed. There were no large differences among the three sectors based on other assessment that take into consideration

students' performances. Participants from the public sector commented that they take into consideration features as class attendance, individual or group participation, homework, presentations and oral performances. Also, participants from this sector (n=17) mentioned that behaviour was another essential aspect during uploading students' marks. A total of 41 participants highlighted that some students might have external difficulties during the class, but they can improve with other alternatives during the course; motivation from students can sometimes be an aspect to help students with their marks.

The private sector showed similarities from the public sector, but they were more emphasis on the students' performance rather than their progress. Features as attendance and meeting deadlines were more relevant for teachers from this sector. Homework, participation during the class and presentations were other aspects that the private sector considers important towards students' marks. Few participants (n=5) stated that they use language portfolios where they can see students' progress in detailed. Language institutes demonstrated that the priority is how students performed in certain assessment and based on the type of assessments teachers can grant another opportunity to redo a task so as to potentially improve students' scores. The use of language portfolios, language autonomy and participations were other factors apart from tests, to complement students' marks.

This section showed the findings from what participants use in their class to assess their students, excluding language tests. Language institutes tend to have more control over their learners in terms of how students progressed during the course by collecting evidence towards students' portfolios, whereas the public sector lacks tracking students' progress, but focuses on how students develop their language and attitudes.

#### **4.2.4 Ethical Considerations in Language Assessments**

Only 141 responses were captured for the analysis. Participants from the public sector recorded the smallest number of entries with 32. The majority (n=28) indicated that there was currently no code of ethics or practise for language assessment purposes. The rest of the participants (n=4) mentioned that in their institutions there were some assessment briefings on how to assess their students, but these regulations were issued by the Ministry of Education (SEP) and there were no updated.

For the private sector obtained 64 statements indicating that participants needed to follow the regulations from the language department in terms of test administrations. Participants (n=3) mentioned that if students did not attend the test-day they were required to pay a fee as a result to take the tests. Two participants mentioned that they followed the methods of regulating

assessment practices and how to deal with areas of cheating during tests and other inaccuracies when assessment takes place, however, this was from the national standards by the Ministry of Education, similarly as the public sector. For the language institutes, 45 entries were recorded, some participants (n=18) indicated that the regulations for taking tests in their institutions were highly regulated by the language programme managers and directors of the institution. Five participants wrote that their school authorities tend to control the assessment practices by developing tests, collecting scores, informing about tests issues and updating other teachers with relevant information about tests.

Reporting a lack of ethical codes and practices in relation to how assessments should be conducted, it could lead to misused and malpractice of the development, administration and communication of results. The following section reports on how participants' experiences in assessments have influenced in making educational decisions as language teachers.

#### **4.2.5 Teachers' Previous and Current Experiences in Assessment**

A total of 201 responses were collected and scrutinised for this theme. Participants mentioned that the current methods they used for their students' assessments are the one they mostly take in their classes with the course books. Exercise, tasks and practical projects from the book tend to be similar since students are familiarised and they find it easier to be completed. Other factor that emerged from this question, was that some teacher (n=19) stated that they use those materials, methods and examples that had an impact on their personal language learning trajectories, and therefore, these teachers continue using these methods since they believe these will have the same impact on their students.

Participants from the public sector also stated that they also find some limitations whilst conducting tests and other testing resources due to a large number of students in their classroom. They administer tasks and exercises that are easy to be evaluated by the teachers. The private sector had similar responses from the public, highlighting that learning a language needs to take practical assessments to empower knowledge and practice based on what students have been learning. The language institute sector has a variety of responses from the other two sectors, the public and private, but they highlighted that they assess how they were previously assessed since they belong to similar institutions.

#### **4.2.6 Critical Assessments Problems in Institutions**

A total of 276 analysed from participants' responses with similar features from the three educational sectors. Nevertheless, the public and the private sectors reported similar critical issues in relation to the lack of constructive feedback to students (n=154). Another similar issue

found in both sectors was the way in which teachers assess students' homework and class participation (n=73). The private sector and language institutes pointed out that the institutional assessment policies were limited, and teachers were not considered in the process of test development (n=112). However, participants from the public sector showed their discomfort from their institutional authorities due to the lack of support with teaching resources and testing instruments to be delivered (n=64). Other critical issues participants' students lack English skills in higher levels and participants' confidence with their English language proficiency.

These six themes reported from Section 2 of TALDI demonstrated that participants' educational sectors treated assessment differently and followed diverse assessment policies. Differences in assessment experiences, practices and policies were found across the public, the private sectors and language institutes. The following section of TALDI (Section 3) was related to how participants' perceived second language testing and assessment backgrounds and their interest towards this educational aspect.

### 4.3 Second Language Testing and Assessment Backgrounds

These findings were reported based on participants' interests and confidence in language testing and assessment (TALDI: Section 3). The first aspect of this section was to identify participants' interests in language testing and assessment in EFL teachers in Tamaulipas. A total of 79 participants from the public sector (72.2%) indicated that they were interested in language testing and assessment; whereas 29 participants (24.8%) from the same cluster pointed out they were not interested. For the private sector, 80.6% participants (n=87) mentioned they were interested in this field, but 21 participants (19.4%) indicated the contrary. For the next cohort, language institute participants obtained a higher percentage of positive interest towards language assessment with 91.2% (n=93) and the lowest percentage of lack of interest with only 9 participants. Participants from the public sector seemed slightly less interested in enhancing their language assessment literacy.

Once participants' interests were analysed, it was essential to identify their perceptions in relation to how prepared or unprepared they perceived themselves towards the development of tests (language testing) and the use of school-based assessments (language assessment). The public sector reported that they perceived themselves as *neither unprepared nor prepared* (n=46) and *somewhat prepared* (n=45) giving a total of 89.7%. Notably, a low number of participants specified that they were *unprepared* for conducting testing practices (N=8), but 6 pointed out they considered *very prepared*. In terms of classroom language assessment, 58 (55.2%) indicated they considered *somewhat prepared* and 30 *very prepared* (28.6%). Sixteen participants were *neither unprepared nor prepared* and only one indicated that is *somewhat unprepared*.

For the private sector, data reported that for language testing 47.2% participants (n=51) considered themselves *somewhat prepared* and 22 participants (20.4%) *very prepared*. The 4.6% of this group (n=5) pointed out they were *unprepared*, and 30 participants indicated that they considered *neither unprepared nor prepared*. By the same token, 63 participants pointed out they felt *somewhat prepared* (58.3%) and 21 *very prepared* for classroom assessment practices. The rest of the group 22.2% felt they were *neither unprepared nor prepared*. From the language institute, participants (3.9%) pointed out they considered somewhat unprepared but the majority, 49 participants (48%) felt *neither unprepared nor prepared*. Thirty-eight participants indicated that they felt *somewhat prepared* (37.3%) and only 11 showed *very prepared* (10.8%). In terms of classroom assessment practices, 50% participants (n=51) considered somewhat prepared and 27 participants *very prepared* (26.5). Nevertheless, 5 participants think they are somewhat prepared and 19 considered neither unprepared nor prepared.

It was also interested to investigate whether the participants had attended any workshop or academic courses in relation to one or various aspects of language testing and assessment. A higher percentage, 84.8%, of participants from the public sector (n=89) had not attended any training in language assessment and 15.2% reported they had attended (n=16). Similarly, the private sector had similar responses, 28 participants reported they had attended, but 80 had not taken any training for language testing and assessment. A slightly similar distribution of participants from the language institute with 56 participants (54.9%) had not taken any previous training, but 46 of them have attended (45.1%). Participants were asked about their interest in a plausible workshop for language testing and assessment so as to improve their quality of testing and assessment practices in the Tamaulipas language classrooms. First, I am reporting on the overall findings in which participants expressed their personal needs regarding to this field. Table 28 presents all the themes which participants considered essential for a training course/ workshop. These were reported on with the means and standard deviation.

Findings showed that participants were mainly interested in *Assessment for language*: receptive skills (4.97), productive skills (4.78), linguistic components (such as assessing pronunciation, spelling and grammar) (4.57), integrated skills (4.33), and the Common European Framework (4.28). The next aspect was the use of *Alternative assessments* such as the use of language portfolios (4.39), peer-assessment (4.29) and self-assessment (4.02). *The use of scales* for measuring language proficiency and performances: analytic scales (4.36) and holistic scales (4.09). *Technical skills* such as development of language tests (4.36), interpreting test-takers' scores (4.19), standardised language tests (4.07), item analysis (4.05), use of statistics (3.96), assessment and testing principles (3.93), methods for language testing (3.91). Then, we have the next category, *assessment as an educational tool*: identify scores to modify teaching practise

## Chapter 4

(4.26), communication of results (3.66). The last category, assessment processes included writing tasks specifications (2.78), purpose for items, tasks and test (3.94), methods for language testing (3.91) and administration of language tests (3.66). an interesting fact here is that participants were not interested in theoretical or historical facts about the field of language testing and assessment.

**Table 28 Overall Participants' Needs for the Workshop**

Assessment theme	Mean	SD
Assessment of receptive skills	4.97	0.16
Assessment of productive skills	4.78	0.47
Assessment of linguistic components	4.57	0.64
Assessment throughout portfolios	4.39	0.92
Development of language tests	4.36	0.66
Use of analytic scales	4.36	0.66
Assessing integrated skills	4.33	0.90
Use of peer assessment	4.29	0.86
Common European Framework of Language Reference	4.28	0.68
Identify scores to modify teaching	4.26	0.68
Interpreting test-takers' scores	4.19	0.69
Use of Holistic scale	4.09	0.86
Standardised tests	4.07	0.73
Item analysis	4.05	0.74
Use of self-assessment	4.02	1.00
Use of statistics for language testing	3.96	1.12
Ethical aspects in testing	3.95	0.77
Purpose for items, tasks and tests	3.94	0.77
Assessment and testing principles	3.93	0.71
Methods for language testing	3.91	0.71
Administration of language tests	3.66	0.83
Communication of results	3.66	0.84
Assessing students with special needs	3.36	1.45
Selection of contents for testing	3.07	0.84
Writing tasks specifications	2.78	0.68
History of testing	1.75	0.70

These findings were vital to analyse what participants considered essential for a workshop, but at the same time these findings also demonstrate participants' current needs in relation to assessment. The following section of TALDI reported the findings on how participants have received professional development support, empowering their professional development for their language courses.

### 4.4 Participants' Continuing Professional Development

The first aspect that data reported was participants' interests in their professional development and their institutional support. The second aspect reported was participants' language and teaching qualifications they had been awarded. Moreover, TALDI also scrutinised participants' responses in relation to what other professional development themes they

considered important for their teaching practice. Lastly, Section 4 of the survey reported the findings on how participants would like to receive training for the further development in a contextualised workshop to empower teachers' language assessment literacy.

Data showed that participants from the three educational sectors were interested in their continuing professional development opportunities. For the public sector, 90 participants (85.7%) were interested and they also indicated that were essential for their workplace. Participants also pointed out that their institutions had been granted support to promote their professionalism with 33 participants (31.4%). By the same token, participants from the private sector also showed similarities in these findings from the public sector, 86 participants (79.6%) were interested and these institutions encourage their teachers to take part in different training and professional activities. In addition to this, 52 participants from this cohort indicated that they had received support from their workplace; whereas 56 participants mentioned that their institution had never provided any aid to empowering teachers' professional trajectories. A discrepancy within this sector, the private, has been reflected among private institutions.

Participants from language institutes reported that 88.2% of its sample was interested in improving their teaching practices, but 12 participants claimed that they had no interest in. Besides this, 75 participants from this cohort mentioned that they had received some support from their institution. With regards to institutional support, they showed they had more possibilities to improve their professional trajectories than the public sector. The three educational sectors shared similar aspects such as funding for conferences or research trips, teaching and language learning qualifications, research leaves, and access to materials and resources.

Participants highest language and teaching qualifications awarded was also examined. From the 315 participants, 12 participants (7 from the public sector, 4 from the private sector and 1 from language institutes) demonstrated that they had not taken any language qualification due to the fact they were bilingual or Mexican-American. The rest of the participants indicated that they had taken TOEFL (n=232) as the highest language qualification, IELTS (n=22) and others (n=49) such as FCE and CAE. For teaching qualifications, all participants indicated they have taken the Teaching Knowledge Test (TKT) Qualification as this was the minimum requirement for teaching in most institutions. The language institute sector (n=46), the public (n=17) and the private (n=4) showed that they had achieved the ICALT and 12 ( 4 public and 8 private) had been awarded CELTA.

A total of 10 CPD themes were included to investigate where language testing and assessment could be placed based on participants' importance to improve their teaching practice.

Table 29 reports on the overall findings in relation to the participants' professional development features. The total number of participants showed that the component of *language assessment* was reported in the sixth position of relevance based on participants' responses. The top five of what they categorised as important for their teaching practice were *the knowledge of the subject, classroom management, teaching materials and resources for language teaching, the use of technology* and *planing lessons*. The features reported with the least means of importance were approaches to language teaching, understanding language learners and educational policies.

**Table 29 Overall Findings of Professional Development Features**

Continuing Professional Development Features	All Participants				Means per Educational Sector		
	Min.	Max.	Mean	SD	Public	Private	L. institute
Knowledge of the Subject	2	5	4.74	0.50	4.95	4.44	4.84
Classroom Management	2	5	4.52	0.88	4.50	4.77	4.28
Materials and Resources for Language Teaching	2	5	4.42	0.80	4.24	4.35	4.67
Use of Technology	2	5	4.35	0.72	4.67	3.91	4.46
Planning Lessons	1	5	4.28	0.93	4.41	3.38	4.65
Language Testing and Assessment	2	5	4.09	1.03	4.82	3.34	4.13
Autonomous Learning	2	5	3.88	1.09	2.93	3.94	4.77
Approaches to Language Teaching	1	5	3.78	1.15	3.40	3.65	4.31
Understanding language learners	2	5	3.68	1.05	3.91	3.11	4.04
Educational Policies	1	5	2.96	0.98	2.92	2.76	3.21

Data has also recorded these findings based on participants' educational sectors. *Language Testing and Assessment* was the first to be analysed. The public sector obtained a mean of 4.82, which ranked in second place of importance; this contrasted the private sector with 3.34 and language institutes with 4.13, both of these were ranked at the eighth place of importance. *Knowledge about the subject* was reported in first place for the public sector and language institutes with means of 4.95 and 4.84 respectively, whereas the private sector scored a mean of 4.44, ranking it in third place of importance. The *Classroom Management* feature obtained 4.50 by the public sector, ranked in third place of importance, with the private sector ranked as the first element of importance with a mean of 4.77 and language institutes (4.28) at seventh place.

For *Planning Lessons*, the public sector captured a mean of 4.41 and language institutes 4.65, both ranked in fourth place. The private sector recorded a mean of 3.81 and was ranked sixth place of importance. *Materials and Resources for Teaching*, this was an important feature as teachers should be able to develop testing and teaching materials. Language institutes achieved a mean of 4.67 ranked at the third place of importance, the fourth place for the private sector with an average of 4.35, and the fifth place of importance for the public sector with 4.24. Regarding *Understanding Language Learners*, the private and the language institute were ranked as the ninth place with means of 3.11 and 4.04 respectively, but the public sector considered this important reaching to the sixth place with a mean of 3.91.

The *Use of Technology in Language Teaching* was in second place of importance for the public sector with a mean of 4.67, language institutes on the fifth place with 4.46 and the private sector in the seventh place with 3.91. *Approaches to Language Teaching* was the sixth place for language institutes with 4.31, the private sector with 3.65 and the public with 3.40 at the seventh. The next theme was *Autonomous Learning* being the second important element by participants from the language institute with a mean of 4.77, the fifth for the private sector with a mean of 3.94 and the public sector in the ninth place with 2.93. The last theme *Educational Policies*, in which all three educational sectors considered the least important; the public sector with 2.92, the private sector with 2.76 and 3.21 for language institute participants ranked at tenth place.

These findings assisted in understanding what teachers seek for opportunities to increase certain areas of the teaching practice to enhance their professional development opportunities, but also those elements to identify any current needs. The following section presents the findings from the participants' assessment skills and practices examined through statistical tests ANOVA.

#### **4.5 Teachers' Assessment Skills and Practices**

This section presents the findings from the TALDI (Section 5) which captured participants' self-perceived assessment skills and practices and has been divided into three parts. The first part of this sub-section reports on the overall finding from the eleven standards of language assessment literacy development. Then, the second part reports on the ANOVA results based on the skills and practices from these previous standards. Finally, it presents the differences among the three sectors: the public, the private and the language institutes.

Table 30 presents the overall findings of participants' self-perceived assessment skills and practices per each standards. The minimum and maximum values are also reported with the total scores that participants captured. This table has two functions, the first one is to report on the descriptive statistics from the assessments skills and the second from assessment practices. The eleven language assessment literacy development standards are presented with their abbreviations in brackets.

**Table 30 Overall Findings from Participants Assessment Skills and Practices**

Language Assessment Literacy Development Standards	Assessment Skills				Assessment Practices			
	MIN	MAX	M	SD	MIN	MAX	M	SD
<b>Knowledge of Assessment</b>	<b>53</b>	<b>60</b>	<b>58.81</b>	<b>1.37</b>	<b>50</b>	<b>60</b>	<b>58.23</b>	<b>1.74</b>
<i>Applied Linguistics (AL)</i>	15	20	19.52	.763	16	20	19.34	.893
<i>Educational Theories (ET)</i>	16	20	19.68	.625	17	20	19.45	.765
<i>Institutional Contexts (IC)</i>	16	20	19.62	.687	15	20	19.44	.848
<b>Technical considerations in Assessment</b>	<b>81</b>	<b>97</b>	<b>89.92</b>	<b>3.24</b>	<b>69</b>	<b>87</b>	<b>78.70</b>	<b>3.61</b>
<i>Testing Principles (TP)</i>	21	25	23.73	.994	18	25	21.63	1.53
<i>Assessment Purposes (APu)</i>	18	25	23.65	1.38	18	25	21.22	1.72
<i>Assessment Processes (APr)</i>	18	25	21.42	1.41	14	21	18.25	1.32
<i>Educational Measurement (EM)</i>	17	25	21.12	1.36	14	21	17.61	1.30
<b>School-based Assessments</b>	<b>98</b>	<b>110</b>	<b>106.8</b>	<b>1.84</b>	<b>82</b>	<b>108</b>	<b>99.36</b>	<b>4.62</b>
<i>Language Assessment (LA)</i>	33	35	34.80	.439	25	35	32.55	2.17
<i>Assessment for Learning (AfL)</i>	19	25	24.03	1.04	16	25	23.35	1.79
<i>Communication of Results (CR)</i>	21	25	24.38	.787	18	25	22.82	1.58
<i>Assessment as Educational Support (AES)</i>	20	25	23.59	1.09	13	25	20.65	1.97

The Assessment Knowledge Component included three standards (*Applied Linguistics*, *Educational Theories* and *Institutional Context*). The overall findings reported based on their mean in assessment skills (M=58.81) and in practices (M=58.23). These two variables scored a positive and significant correlation ( $r = .67$ ,  $n=315$ ,  $p < .001$ ). These three standards had a total of 4 items maximum; therefore, with a Likert-scale from 1 to 5, the maximum value for the mean was 20 in each giving a maximum of 60. Firstly, for the standard of Applied Linguistics (AL), participants scored a mean of 19.52 for assessment skills, a slightly lower mean was captured for assessment practice (M=19.34). The relationship of the perceived assessment skills and practices from this standard was examined using Pearson product-correlation coefficient. There was a small positive correlation between these two variables ( $r = .16$ ,  $n=315$ ,  $p < .001$ ). Then, the educational theories (ET) in which participants obtained a mean of 16.68 in assessment skills and in practices a mean of 19.45. These variables obtained a strong positive correlation ( $r = .51$ ,  $n=315$ ,  $p < .01$ ) with higher levels of perceived assessment skills and practices. The standard of *institutional context* (IC) reported a mean of 19.62 in self-perceived assessment skills and a mean of 19.44 in practices. These variables obtained a significant positive correlation ( $r = .59$ ,  $n=315$ ,  $p < .01$ ).

The next Component was Technical Practices that includes four language assessment literacy development standards: testing principles (TP), Assessment purposes (APu), Assessment processes (APr) and Educational Measurement (EM). The overall findings reported based on their mean in assessment skills (M=89.92) and in practices (M=78.70). These two variables scored a positive and significant correlation ( $r = .45$ ,  $n=315$ ,  $p < .001$ ). These standards had a total of five items which indicates that the maximum value was 25. Testing principles obtained a mean of

23.73 in self-perceived assessment skills and practices with 21.63. These two variables scored a positive and significant correlation ( $r = .37$ ,  $n=315$ ,  $p < .01$ ). Assessment purposes gathered a mean of 23.65 in assessment skills and 18.35 in practices. These two variables obtained a small, but significant correlation ( $r = .34$ ,  $n=315$ ,  $p < .01$ ). Assessment processes obtained a mean in skills ( $M = 21.42$ ) and in practice ( $M=18.25$ ) and these two variables obtained a lower positive correlation, but not significant ( $r = .10$ ,  $n=315$ ,  $p > .01$ ). Educational measurement, similarly, as (Apr), this component obtained a mean of 21.12 in skills and 17.61 in practices. Pearson correlation was conducted, and it was reported with a small positive correlation, but with no significance ( $r = .05$ ,  $n=315$ ,  $p > .01$ ).

The last component was Classroom practices, a total of four standards fell into this component: Language Assessment (LA), Assessment for learning (AfL), Communication of results (CR), and Assessment as educational support (AES). The overall findings reported based on its mean in assessment skills ( $M=106.8$ ) and in practices ( $M=99.36$ ). These two variables scored a small, positive and significant correlation ( $r = .24$ ,  $n=315$ ,  $p < .001$ ). Language assessment standard had seven items, whereas the rest of the standards in this component had five items. Language Assessment obtained a mean in participants' self-perceived assessment skills of 34.80 and in practices 32.55. These two variables were examined with Pearson correlation and this reported on with a small positive correlation but significant ( $r = .15$ ,  $n=315$ ,  $p < .01$ ). Assessment for learning obtained a mean in assessment practices of 24.03 and in practices 23.35. These variables scored a small positive correlation and significant ( $r = .22$ ,  $n=315$ ,  $p < .01$ ). The next standard was communication of results, in which participants obtained a mean of 24.38 in assessment skills and 22.82 in practices; this was a small positive correlation captured and it was not significant ( $r = .04$ ,  $n=315$ ,  $p = .441$ ). Lastly, the last standard was assessment as an educational support with a mean observed ( $M=23.59$ ) from participants' self-perceived assessment skills and practices ( $M=20.65$ ) with ( $r = .03$ ,  $n=315$ ,  $p = .959$ ).

The next part of the analyses was to find any discrepancy of participants' self-perceived assessment skills and practices. A Multivariate Analysis of variances (MANOVA) was conducted to identify differences among the participants' educational sectors. For the purpose of this analysed two hypotheses were stabled. The null hypothesis indicates that participants across the educational sectors do not show any differences in their skills and practice. The alternative hypotheses suggest that participants show differences in their skills and practices. These findings are reported on the language assessment literacy development standards across the educational sectors, and then they are presented with the analysis of each of the components: knowledge of assessment, technical considerations and school-based assessment s.

Table 31 reports the multiple comparisons of the dependent and independent variables. It illustrates the difference among the educational sectors, differences of means in groups and standard error, as well as their significance. There were some differences in means found among participants' educational sector. In the first Component (Knowledge) significant difference was captured based on Tukey's analysis. In relation to self-perceived assessment skills and practices of the component of knowledge, participants from language institutes indicated that there was a statistical difference with the public and private sector participants skills and practices in the component of knowledge ( $p = .000$ ). No statistical difference was found between the public and private sector skills ( $p = .104$ ) and practices ( $p = .620$ ).

**Table 31 Multiple Comparisons among Participants' Educational Sectors**

Dependent Variable	Educational sectors		D.	Std. Error	Sig.
	Public	Private			
Knowledge of Assessment (Skills)	Public	Private	.36	.177	.104
		Language Institutes	<b>-.79*</b>	.179	.000
	Private	Public	-.36	.177	.104
		Language Institutes	<b>-1.16*</b>	.178	.000
Knowledge of Assessment (Practices)	Public	Private	.21	.227	.620
		Language Institutes	<b>-1.12*</b>	.230	.000
	Private	Public	-.21	.227	.620
		Language Institutes	<b>-1.33*</b>	.228	.000
Technical considerations in Assessments (Skills)	Public	Private	<b>3.40*</b>	.347	.000
		Language Institutes	<b>-1.46*</b>	.352	.000
	Private	Public	<b>-3.40*</b>	.347	.000
		Language Institutes	<b>-4.86*</b>	.349	.000
Technical considerations in Assessments (practices)	Public	Private	<b>5.81*</b>	.365	.000
		Language Institutes	<b>1.82*</b>	.370	.000
	Private	Public	<b>-5.81*</b>	.365	.000
		Language Institutes	<b>-3.99*</b>	.367	.000
School-based Assessments (Skills)	Public	Private	<b>1.48*</b>	.237	.000
		Language Institutes	.27	.241	.495
	Private	Public	<b>-1.48*</b>	.237	.000
		Language Institutes	<b>-1.21*</b>	.239	.000
School-based assessment (Practices)	Public	Private	<b>6.28*</b>	.480	.000
		Language Institutes	-.20	.487	.912
	Private	Public	<b>-6.28*</b>	.480	.000
		Language Institutes	<b>-6.48*</b>	.484	.000

The second component was technical considerations in assessment. The findings from this component reported statistical differences in participants' assessment skills and practices across the sectors ( $p = .000$ ). Therefore, this indicated that the participants' assessment skill and practices in this component were different and the null hypothesis must be rejected. In relation to school-based assessment, participants' skills and practices were statistically different. The null hypothesis was rejected between the public sector and language institutes in participants skills ( $p = .495$ ). The alternative hypothesis stated that the groups of participants' assessment skills and

practices matrices of variances were not equal. Cohorts from the private and language institutes reported statistically differences in their classroom assessment skills ( $p = .000$ ). Regarding school-based assessment practices, the public and language institutes did not reflect difference ( $p = .912$ ) but significant difference within this component was captured with the private and language institutes ( $p = .000$ ).

The next part of the analysis was to report if there was any significant difference among the three educational sectors and participants' self-perceived skills and practices under each standard of the language assessment literacy development framework. In relation to the first standard, Applied Linguistics, data reported that participants' skills had no significant difference between the groups of the public sector and language institutes ( $p = .142$ ), but significant difference was reported on the public and private sectors ( $p = .002$ ) and language institutes with the private sector ( $p = .000$ ). Participants' practices under this standard, Applied Linguistics, showed that there was no significant difference between the groups of the public sector and language institutes ( $p = .418$ ) but significant differences were reported in the public and private cohorts and the private sector and language institutes ( $p = .000$ ).

In the second standard, Educational Theories, data reported that participants' skills had no significant difference between the groups of the public and private sectors ( $p = .885$ ), but significant differences were reported on the public and language institutes ( $p = .002$ ) and private and language institutes ( $p = .009$ ). Participants' practices under this standard demonstrated there was no significant difference between the groups of the public and private sectors ( $p = .956$ ), but significant difference was reported on language institutes and public, as well as the private sector ( $p = .000$ ).

Participants' skills of the third standard, Institutional Context, reported that participants' skills had no significant difference between the groups of the public and private sectors ( $p = .782$ ), but significant differences were reported on the public and language institutes ( $p = .003$ ) and the private and language institutes ( $p = .000$ ). Participants' practices under this standard demonstrated there was no significant difference between the groups of the public and private sectors ( $p = .135$ ), but significant difference was reported on public and language institutes ( $p = .000$ ) and language institutes with the public sector ( $p = .024$ ).

In the fourth standard, Testing Principles, data reported that participants' skills had significant difference across the three groups obtaining a significant difference ( $p = .000$ ). Similarly, participants' practices under this standard demonstrated there was significant difference between the groups of the public and private sectors ( $p = .024$ ), and private sector and language institutes ( $p = .000$ ). Data from participants' skills, in Assessment Purposes, reported that

## Chapter 4

participants had no significant difference between the groups of the public sector and language institutes ( $p = .296$ ), but significant differences were reported on the public and private sector and private and language institutes ( $p = .000$ ). Participants' practices under this standard demonstrated there was a significant difference among the three educational sectors ( $p = .000$ ).

Significant differences were also reported among the three groups of participants' skills in the standard of Assessment processes ( $p = .000$ ). Likewise, participants' practices in this standard had significant differences: the public and language institutes recorded ( $p = .046$ ) and language institutes and the private sector ( $p = .007$ ), the public and private sector ( $p = .000$ ). The seventh standard, Educational Measurement, recorded a significant difference on participants' skills from the private and language institutes ( $p = .000$ ), but no significant differences were reported within the private and public sectors ( $p = .143$ ) and language institutes with the public sector ( $p = .143$ ). In relation to participants' practices of this standard, significant differences were recorded across the groups ( $p = .000$ ).

The next standard, Language Assessment, reported a significant difference within the groups of the public and private sector ( $p = .000$ ), no significant differences were recorded with the public and language institutes ( $p = .052$ ) and the private sector and language institutes ( $p = .285$ ). Regarding participants' practices, statistical differences were found in the public and private sector, as well as language institutes and the private sector ( $p = .000$ ). The following standard, Assessment for Learning, recorded significant differences in participants' skills of from the private and public ( $p = .015$ ), the public with the language institute ( $p = .009$ ) and language institute with the private sector ( $p = .000$ ). For participants' practices, no significant difference was reported between the public and language institutes ( $p = .097$ ) but a significant difference was recorded with the private sector and language institutes ( $p = .000$ ). Data reported that there was no differences among the participants' skills in relation to the standard of Communication of results; the public and private sector ( $p = .769$ ), private and language institutes ( $p = .643$ ) and the public sector with language institutes ( $p = .979$ ). By the same token, for participants' practices, no significant difference was recorded within the public sector and language institutes ( $p = .209$ ) and the private sector with language institutes ( $p = .058$ ). Significance difference in participants' practices from the public and private sector was observed ( $p = .000$ ).

Lastly, participants' skills of the standard of Assessment as an Educational Support reported a significant difference between the public and the private sector ( $p = .000$ ) and the public with language institutes ( $p = .000$ ). No difference was captured for participants' skills for the private sector and language institutes ( $p = .246$ ). For participants' practices under this standard showed that significant differences were captured for the public and the private sector ( $p = .000$ ) and the

private sector with language institutes ( $p = .000$ ). No significant difference was reported between the public sector and language institutes ( $p = .415$ ).

#### 4.6 Summary of the Chapter

The purpose of this Chapter was to report on the findings from the survey entitled Teachers' Assessment Literacy Development Inventory. A total of 315 participants completed the survey: public sector ( $n=105$ ), the private sector ( $n=108$ ) and language institutes ( $n=102$ ). Some of the most relevant findings from the analysis of the five sections of TALDI data were:

- In terms of the academic profiles of the participants, the private sector have obtained higher academic qualifications such as diplomas and master's degree compared to the public and language institute sectors. However, language institutes tended to have more academic profiles related to language education. The public sector and language institutes had less participants holding degree from other disciplines than education. The private sector, on the other hand, demonstrated that a large number did not have any academic credentials towards language education.
- The number of students in a Tamaulipas classroom depends primarily on the type of educational sector. Findings showed that participants from the public sector had larger groups of students compared to the private and the language institutes.
- Teachers' working hours was also explore and findings reported that only the public sector teachers tended to work higher number of teaching hours. The private and the language institute indicated that the average of teaching hours is between 16 to 20 hours a week. Similarly, participants from the previous two sectors, also stated that they teach the shortest period between one to five hours a week.
- Students' language proficiency courses varies from each of the sectors as well, participants from the public sector demonstrated that they teach lower levels of English proficiency (A1 and A2); whereas the private and the language institute teach intermediate and advanced proficiency levels. This is since the private and language institutes also offer language certifications to their students.
- Test development responsibilities relied on both; teachers and managers across the sectors. However, the public sector demonstrated that their tests were designed by these teachers with more frequency than the other two sectors (the private and language institutes). Similar number of participants from the private and language institutes do not develop their tests since their programme managers do this activity. Most of the test developed are taken from their teachers' course-books, internet and mock tests. Also, it is important to highlight that some teachers from the three sectors neither develop nor give tests to their students.

## Chapter 4

- Participants from the language institutes tended to demonstrate an ample methods to assess their students' proficiency in relation to the four language skills and grammar and vocabulary. The public sector had the majority of their assessments towards understanding and applying grammar and vocabulary, whilst the private sector demonstrated to assess mainly receptive skills (reading and listening).
- The use of alternative assessments (e.g., portfolios, presentations, etc.) among participants from the three sectors did not show significant differences, but there were other factors that teachers take into consideration for assessment purposes such as attendance, participation and homework (affective factors).
- Ethical considerations from the administration of the language tests were neither found in the public nor the private sectors, but they have some administrative files in which they detail how assessments should be administered to students. However, language institutes have adequate briefings to conduct and administer language tests, stipulated and designed by their programme managers.
- Some critical issues of assessments found among the three sectors were towards methods for delivering relevant feedback to students, the assessment of students' participation during the class and homework. In addition, a factor was found in participants responses from the public sector in relation to teachers' lower English proficiency level and their confidence.
- Participants demonstrated that they were interested in the field of language assessment. However, there was a small number of participants who perceived themselves as unprepared in the field of testing and assessment. Mertler (2009) suggested that sometimes teachers underestimate their potential when they had to answer how skilled or prepared are in certain disciplines.
- Participants had different perspectives on the use and application of assessment. Based on participants' interests in this field, the most attentive themes in language testing and assessment from teachers in Tamaulipas were assessment of the receptive language skills (reading and listening), the productive language skills (writing and speaking), and linguistic components (grammar and vocabulary), assessment through language portfolios, development of language tests, use of the scales and the Common European Framework. The least interested themes were history of language testing, writing task specifications, selecting contents for testing, assessing students with special needs and communicating assessment results.
- Professional development was also an essential factor that was reported as participants indicated that they were interested in their continuing professional development, but the majority of the private and language institute sectors had received more support from their

institutions. Being the public sector that has less opportunities to improve their professionalism towards the field of EFL.

- Each institution has its own policies at the moment of hiring their teachers. Some institutions have less rigid methods such as the public sector; the private on the other hand, required more experience towards the expertise and use of the language. The language institutes tended to be the most rigid and cautious at the moment of hiring personal. Participants from the three sectors indicated that they have awarded TOEFL, ICELT, FCE and CAE as language qualifications and for teaching qualifications, all participants pointed out they have achieved TKT.
- Towards professional development, participants considered other areas more relevant than assessment. Language assessment was not a priority for continuing professional development for all the population, but for the public sector was considered the second aspect to empower through professional development.

Language institutes demonstrate to have differences among the public and the private sectors in the component of knowledge of assessment skills and practices. A significant finding was the statistical differences between TLASP from the technical considerations in assessment among the three sectors. It is worth noting that more discrepancies, among these sectors, may rise when analysing the findings from the interviews, since these findings require to triangulated with qualitative data. Finally, no differences were reported with the component of school-based assessment with the language institutes and the public, but differences were reported on the private sector.



## Chapter 5 Semi-structured Interview Findings

This Chapter reports on the qualitative findings from the semi-structured interviews these findings came from the mixed-method approach for the exploratory research. This analysis was crucial to answer research questions related to the status quo of Tamaulipas EFL teachers' and language programme managers' language assessment literacy, the differences between educational sectors and their professional development trajectories. Seven themes emerged from the data collected: 1) language tests (relevance and occurrences), 2) language test development: language skills and linguistic components, 3) backgrounds in professional development and language and assessment, 4) teachers' previous and current assessment experiences, 5) teachers' assessment knowledge and technical considerations, 6) teachers' school-based assessments, and 7) critical assessment issues among educational sectors. This Chapter concludes with a summary of the most relevant findings of Phase 2 (interviews) for the exploratory research.

### 5.1 Language Tests (Relevance and Occurrences)

This section presents the findings from the use of tests in education, their importance and how often EFL teachers in Tamaulipas administer these tools to their language students. All participants (n=15) stated that *language tests*, as well as '*language examinations*' (standardised tests), were used in their institutions. Some of the test purposes were to make institutional judgments, deciding to pass or fail students based on their language proficiency and to measure students' progress trajectories. Participants across the sectors also indicated that tests were for both formative and summative assessments. Only two participants from language institutes declared the use of diagnostic tests at the beginning of their English language courses.

Participants from the public sector and language institutes specified that tests were necessary for their language courses because these instruments could provide a significant number of educational benefits, not only for the students, but also for the teachers. These participants from the public sector pointed out that tests were aligned to the course contents. Participants from the private sector and the language programme manager pointed out that the tests were used as guidelines to observe students' progress, but also they stated that sometimes tests should not be the only method to demonstrate whether the learning objectives of a course were met. Participants from the private sector also highlighted that without the use of tests for educational purposes, they could spend more time observing and capturing students' achievements. Similarly, a participant from the language institute opined that tests were also resources to observe what and how students have learnt in a particular period, reflecting more towards the use of tests for educational support. All language programme managers from the

three sectors shared similar thoughts in relation to the importance of language tests benefiting both teachers and students.

*'When I mark the tests, I can identify my students' strength and weaknesses... I can provide written feedback' ([INT] LI-T3).*

*'Tests and quizzes are great sources to identify how students have been learning and what they require to learn further' ([INT] PUB-T2)*

*'I use tests to identify what I need to teach better so as to students can get better marks next time' ([INT] PRIV-T3)*

PUB-T1 indicated that the use of tests in the classrooms could allow students to understand and gain experience in preparing them for 'official exams': in other words, standardised summative tests. Also, this participant considered that by practicing these tools, students could be more prepared to a diverse selection of tests for students' future testing opportunities. The language programme managers from the private sector and the language institutes stated that international language tests such as the Cambridge Main Suite and TOEFL were the most commonly used by the students at their workplace. They expressed their approval for these examinations due to the high standards and international recognition these language qualifications could offer. The language programme manager from the private sector also reported the used of standardised tests to guide institutional authorities regarding the performance of teachers and students, but not for the purpose of administering these to students as summative assessments. It seems that some institutions administer these standardise tools to also track teachers' practice. However, participants from the language institutes expressed that standardised tests were used with the only intention to assess (summative) their students 'language proficiency' at the end of their language learning courses, especially higher language courses.

*'Tests are the best tools to prepare and train students to take official tests like TOEFL or FCE' ([INT] PUB-T1).*

*'We used Cambridge Exams at higher levels so students can understand how much they have learnt and what teachers should prepare for their classes' ([INT] LPM-LI)*

All participants were also asked about how often they administered testing resources and the number of tests they considered was enough and acceptable for learning purposes. Although all participants verbalised a different number of tests occurrences, the average of the public sector, as with the private sector, was 3 times (reporting up to 4 for higher education institutions) and language institutes averaging 4 times per term (within a period of six months). Several participants indicated that the number of tests relied on institutional regulations and the number of teaching hours. A relevant finding was that three participants from the public sector and one from the private sector mentioned they could negotiate the number of test administrations with

their students. These participants mentioned the accessibility between teachers and students to define the assessment methods. The majority of participants across all sectors pointed out that tests were usually administered after teachers had already taught the contents that needed to be tested, according to the institutional policies or negotiations from their students. Some emerging reasons behind participants' decisions relating to the number of tests were that these facilitated the understanding of students' performances; how well or badly their students have been progressing, as well as how teachers could report their students' performances into marks or grades, and tracking their learning pathways across the language skills.

Findings from this sub-section reported that participants had similar thoughts in terms of the benefits and purposes of tests and standardised testing instruments. EFL teachers and programme managers mainly relied on language tests and were conscious of different types of tests and forms. Standardised tests were frequently discussed by the private and the language institute sectors, whereas for the public sector, these tools seemed to be used with less frequency. The number of tests administered in each educational sector varied based on the contents of the curriculum and the hours of instruction. Differences were reported from the public and private sector where upper-secondary schools tended to have a smaller number of tests per term than the private sector. The following section presents the findings on participants' educational sectors in relation to test development and which aspects of language they usually assessed their students.

## **5.2 Test Development: Language Skills and Linguistic Components**

The second theme focuses on how teachers develop their tests, if they do not develop them, then who develops them, as well as what language skills and structure features teachers take into consideration to develop their testing instruments. All participants from the three sectors indicated that they used different methods to develop their tests. Three participants from the public sector indicated they were required to develop their own tests. The other participant from this cohort indicated that tests were given by the language programme manager. PUB-T1 pointed out that the test contents were taken from the course textbooks and online resources. Another example was from PUB-T2 and PUB-T4, who indicated they had adapted the contents from the National Curriculum of English for Basic Education, developed by the Ministry of Education. A discrepancy from participants of this sector was identified based on how teachers can have access to different testing materials and responsibilities.

From the private sector, only one developed the tests. This one participant (PRIV-T3) commented that the development of these tests was through extracting items from course textbooks or software; a similar response from the participant from the public sector (PUB-T1).

PRIV-T1, T2 and T4 claimed that the instruments developed by their language programme managers may not always meet students' current needs and that this could possibly lead to problematic scenarios where assessment becomes unfair and biased. Although participants from this cohort were able to develop some (alternative) assessments, these did not have a significant impact towards students' summative assessments.

*'I use the National Programme of English for Basic Education since it is a complete version of what students should know and do in the English classrooms in Mexico, as well in here in Tamaulipas' ([INT] PUB-T2)*

*'We deliver the tests based on the contents from the course but regardless of this, students still getting low marks. I believe that it is when assessment takes place and some teachers are not well-prepared to teach a language' ([INT] LPM-PRIV).*

All participants from the language institutes claimed that their language tests were developed and provided by their language programme managers. A participant from a language institute pointed out that the contents of the tests were similar to the ones taught in the classroom and mostly aligned to the course objectives. The language programme manager from the private sector also indicated that the tests included contents that teachers and learners were currently revising during their lessons. Two participants (PRIV-T1 and LI-T3) said that the development of tests was restricted to the teachers in their institutions. They indicated that teachers from their institutions were not considered to assist or even revise the process of test development. Contrasting these previous participants, one participant claimed that these tests were highly reliable, but these instruments were not designed for a specific group of students: in fact, not to all students from the same language proficiency level. Participants, mainly from language institutes, felt the need to acquire certain "control" for developing tests and materials since the teachers were the ones who knew the students' needs and their progress.

*'Teachers have no access to develop or mention anything related to language tests, they are created by the coordinator and we have to administer them' ([INT] PRIV-T1).*

*'I believe tests are created based on the contents because they rely on the units I teach but these are not clear and sometimes they include contents that were not even discussed' ([INT] LI-T2).*

An interesting factor to analyse was the structure of participants' tests per educational sector. The public sector mentioned that their tests were developed mainly to assess students' ability towards grammar and vocabulary. Two participants stated that they had developed several reading and listening tests. PUB-T2 and PUB-T4 pointed out that due to a lack of time they preferred to test grammar with certain exercises and occasional listening tasks, as these instruments were easier to mark and provided accurate and timely feedback to their students. This difference was seen within the upper-secondary schools' participants, as well as with higher education in the public sector. This finding could be the reason why the public sector tends to

have a significant number of students in each classroom. These findings could also suggest that participants may be equipped with basic knowledge by identifying the assessment purposes and deal with different situations in their institutions.

Three participants from the private sector indicated that they mainly assessed three language skills (reading, listening and writing) through tests. In contrast, PRIV-T2 also assessed linguistic components (e.g., grammar, vocabulary, coherence, etc...) embedded with reading and writing skills (integrated skills). The language programme manager from the private sector claimed that the assessment for grammar and vocabulary relied mainly on the students' written performances. Participants from the private sector mentioned that they tended to assess speaking skills in different forms, such as classroom participation or group presentations; but these were less formal than a language test. This indicated that the private sector included integrated language skills to assess their students' communicative ability of language use.

Participants from the language institutes showed that they assessed all four language skills. Also, these participants said that in their tests, there was a specific section for grammar and vocabulary, known as 'Use of English' and 'Written Expressions'. LI-T4 pointed out that their tests included all language skills, but that the contents relied on the topics that were taught by the teachers. In addition, participants from this sector were the only ones who also indicated that the use of the CEFLR was taken into consideration for the language tests. Similarly, the programme manager indicated that teachers were required to understand this framework in order to be able to assess students' performances mainly for the productive skills, and to teach the language appropriately.

*'The tests include all the contents from the classes and the tests are well designed by the coordinators that include most of the contents mainly relying on the Common European Framework to develop the test and choose the contents' ([INT] LI-T4).*

This sub-section reported that participants from the public sector tended to understand better their testing resources due to the fact they developed their own instruments. Although they developed their tests, these instruments could not necessarily test what they were supposed to measure. There were some limitations presented regarding the public sector as they did not have equal opportunities for testing, mainly for productive skills, whereas the private sector indicated that whilst writing was frequently assessed, speaking was discarded in language tests. Participants from the private sector and language institutes mainly received the tests by their programme managers, as they had restrictions in developing their own tests due to the required institutional standards. The following section discusses the findings from participants' continuing professional development and their language assessment backgrounds.

### 5.3 Continuing Professional Development and Language Assessment

This sub-section has been divided into three parts based on the emerging themes of the qualitative analysis: 1) participants' perceptions towards professional development, as well as, trainings and qualifications, 2) professional development to enhance language assessment literacy, and finally, 3) a lack of professional development for language testing and assessment.

#### 5.3.1 Professional Development: Trainings and Qualifications

All participants considered that professional development was important for their performances as language teachers. They showed interests in different professional development components such as: planning lessons, development of materials and resources for learning, understanding their learners, the use of technology and autonomy, but none of the teachers nor language programme managers across the three different educational sectors mentioned assessment as a priority area of professionalism.

The manager from the private sector defined professional development as an expansion of teachers' academic trajectories, which connects *'teaching practices to reflection'*. By the same token, this pointed out that professional development was *'extremely'* important since it relied on the foundations to become a good teacher. The manager from the language institute also opined that professional development was a form to reflect, analyse and to gain new skills and practices to conduct inside the language classroom. Regarding the public sector, the manager said that professional development was the practice of working in collaboration with other teachers to improve the quality of instruction within an institution. The three language programme managers had in common that the continuing professional development is vital to boost teachers' skills and practice for the appropriate application of these in the language classroom.

*'Professional development is an opportunity to reflect on our own practice and think how we can improve'* ([INT] LPM-LI)

*'It is extremely important to share ideas with other teachers so as to improve the quality of instruction. To gain and practice new approaches to expand our knowledge and techniques'* ([INT] LPM-PUB)

Participants from the public sector claimed that they required to take different professional development activities such as: teacher trainings, mentoring courses, observations and workshops with other colleagues from the same institution as part of their job duties. Similarly, as participants from the public sector, two teachers from language institutes also exposed their interest in attending general and specialised courses in order to be promoted to higher positions and enhanced their teaching competence.

Likewise, three teachers from language institutes specified that their institutions motivated their teachers to attend different trainings, but some of these professional development opportunities were mandatory to empower their teaching skills. On the other hand, one participant from the private sector brought up that teachers' professional development should be essential to identify each individual's strengths and weaknesses, but should not be compulsory. This participant (PRIV-T3) also indicated that some institutions require teachers to attend a series of seminars that this private institution had developed for its employees. Three participants, two private sectors and one from the public, consider their teaching qualifications as professional development opportunities and reflected that these were more valuable than other activities like workshops, seminars or courses.

Participants were also asked whether they had been awarded any teaching qualifications and the impact of these qualifications on their professional development as language teachers. The three programme managers stated that their teachers were required to demonstrate three aspects in order to be hired in their institutions: 1) relevant educational background, 2) language proficiency qualifications, and 3) teaching qualifications. However, the public manager indicated that the process for hiring new personnel utterly relied on the National Educational Policies: in other words, external to the institution. The public sector and language institute managers said that a bachelor degree (B.A) in relation to language teaching pathway was essential to have, and international qualifications that could demonstrate their teachers' knowledge, practices and skills in language education. These findings contrasted from the one in the private sector, who mentioned that the degree in language teaching was desirable.

Participants from all three sectors had been awarded with one or more teaching qualifications, as well as a language qualification being the minimum level of Proficiency B2 (CEFR). However, two participants from the public sector confessed they only have been awarded with a lower level B1 (CEFR) language qualification. Participants from language institutes said that the TKT and ICALT were compulsory for the purpose of promotion and monetary increase. Similarly, two participants from the public sector communicated that they had to take certain teaching qualifications as part of their work since these were imposed by the government: CENNI and PNIEB. These previous are from the National Framework for Languages.

*'Seminars, courses and conferences can help us understand how to deal with different aspects of teaching and how to improve our standards' ([INT] PRIV-T1)*

*'Development of our professional progress, as teachers, could assist in improving the quality of instruction and empowering with better tools. That is why I have to attend several trainings developed by our Director' ([INT] PRIV-T3).*

Two participants, PRIV-T1 and PRIV-T3, brought up that teacher training courses should be taken into consideration to provide better teaching practices, as well as to understand what the students and institutions currently require. All language institutes participants pointed out that training courses could have a significant impact on their teaching practices. This was reflected based on the compromise between the manager and the teachers to examine what features teachers need to empower the performance of students, teachers and above all, the institution. Data reported that participants had different opinions on how professional development can be applied to empower teaching practices. Also, a discrepancy between training and development was observed within participants across the sectors. The following section presents the findings from professional development towards language assessment literacy.

### 5.3.2 Professional Development in Language Assessment Literacy

In relation to participants' professional development in the area of language assessment literacy, six out of 15 participants, one from each sector and all the programme managers, had previously taken a specialised training course, attended a conference, seminar or a workshop in language testing and assessment. The three managers indicated that they had also taken one or more teaching qualifications in which they had revised different aspects of language assessment, but not in-depth. Two participants from the public sector pointed out that during their studies at their undergraduate programme in language education; they were able to revise briefly some theoretical and practical concepts on how to assess students' language skills. Some of the assessment contents these participants had access during their educational trajectories were different types of assessments (formative and summative), delivering feedback, test development and use of holistic and analytic marking schemes.

*'...when I was doing the ICALT, I attended a session related to assessment and evaluation but this was mainly theoretical and not practical at all' ([INT] PUB-T2)*

*'I have taken many teacher training courses and qualifications, but I believe that the component for assessment was not adequate for my teaching practice' ([INT] PRIV-T3)*

*'I had the opportunity to attend a training session on how to improve students' writing through feedback. This was very useful more than any teaching qualification' ([INT] LI-T4)*

These participants demonstrated that some courses were vital and engaged with current situations. However, participants indicated that there was a lack of contextualised courses that teachers currently need. Seven (all from the public sector, two private and one language institute) participants highlighted the lack of assessment component in most of the professional development courses they have attended. The three language programme managers and four teachers suggested that there should be an assessment training course or module in assessment offered to language teachers in which this could include different skills and practices to be used in

different contexts. A participant commented that language assessment was briefly covered in teacher training courses they have previously taken. The manager from the public sector pointed out that language assessment was not often included in training courses that teachers were required, and that taking and taking extra courses would be difficult. In fact, a participant from the public sector (PUB-T2) also stated that these professional development courses were offered by the Mexican Ministry of Education (SEP), which focused primarily on aspects like teaching the language, rather than assessing it. One participant from the public sector (PUB-T4) stressed the lack of assessment training courses especially when local and national authorities pressured to increase students' scores in this sector. One participant (LI-T4) indicated he had taken specialised modules in assessing English language before as part of professional development activities. This latter participant stated that the focus of this training was on to improve teachers' feedback for assessing students' written performances. This participant also claimed that the methods on how to assess language learners were not appropriate taught in teaching qualifications.

Data informed that participants were not content with the syllabus from international teaching qualifications and they highlighted that the contents should be much more practical than theoretical. In addition, participants' educational backgrounds could help in boosting their professionalism, but this could not be enough to be called assessment literates. Moreover, participants raised their concerns in relation to their language assessment literacy since they considered that the educational policies had not tackle this issue effectively. The following section presents the findings from participants' voices in relation to lack of professional development in developing language assessment literacy, as well other areas of professional development features.

### **5.3.3 Lack of Professional Development in Language Assessment Literacy**

A general finding was that even though all teachers require continuing professional development, reality does not seem to support teachers at all. In this section, several reasons for not attending professional development opportunities are discussed from the emerging data. Most of the participants from the three sectors mentioned that it was due to a *lack of time* because of working hours that does not allow them to take further opportunity to progress professionally. Participants who held a full-time position demonstrated that they were unable to attend further training to develop their assessment literacy, one participant also considered that there was not needed because he had already learn all these concepts from previous teaching courses. Participants from the public sector had in common that are required to work extra hours for marking students' homework, essays, language portfolios and presentations. However, other

reasons participants highlighted were in relation towards participants' family responsibilities and other work-related activities.

Beside this, participants' *lack of interest* was also reported from the qualitative data. Participants considered other features of professional development more relevant and important than language testing and assessment in their classroom. The manager from the public sector indicated that there was a need to improve their assessment methods at institutional level, and it was crucial to train teachers since they were mainly responsible for the assessment practices. Also, two participants from the language institute indicated that teachers should be trained consistently in different areas, assessment was only mentioned twice. In relation to the private sector, participants demonstrated that they were not mainly responsible for the institutional assessment practices; therefore, this feature of professional development was not part of their professional interest.

*'Assessment is an important aspect for my teaching, but I believe there are other factors more important like classroom management or developing materials'* ([INT] PRIV-T1)

Data also reported other emerging reasons why participants might not be interested in taking professional development courses with regards to the relationship between teachers and the field of language assessment. These were a) *the lack of knowledge related to the subject*, b) *financial aspect*, c) *the length of the courses* and d) *teachers' current English proficiency level*. Participants pointed out that the materials and courses available to improve testing and assessment knowledge were complex to understand. These participants claimed that the contents and the methodology of these sources were for experienced teachers and sometimes were not applicable to their teaching contexts. Also, participants tended to relate assessment practices with the use of number, algorithms or statistics. Two participants from the language institute and one from the public sector highlighted that assessment requires a lot of effort to understand the way of marking, not based on a criteria, but on how the criterion could be used. This demonstrated that the lack of prior knowledge and participants' experiences towards assessment might have a negative impact to engage teachers to have access to professional development training to improve teachers' language assessment literacy

The other reason was the lack of financial support from their working institutions. Participants from the public sector indicated that they were constantly trained to be updated with the newest trends in the field of language education with teacher training courses delivered by the Ministry of Education (SEP). The manager from the language institute stated that teachers should be responsible for their own professional development, but unfortunately there was a lack of support from institutions. Although, two participants, PUB-T1 and PRIV-T3 showed their concerns about the *lack of the monetary* support that allows them to continue increasing their

teachers' professional profile in those areas they perceived essential. Programme manager from the private sector pointed out that support is given to several teachers, but there were only for exceptional candidates and who have full-time status.

*'Teachers should attend conferences and seminars to become better in different areas, but attending training events were relatively expensive and their workplace did not necessarily need to provide a kind of support' ([INT] LPM-LI)*

The third reason was commonly found within participants from the public sector in which two stated that the duration of courses and qualifications *was excessively long and required a lot of attention*. Teaching qualifications tend to be long and this also affected them based on *the lack of time*, which participants find difficult to concentrate and to successfully finish this. The manager and a participant from the private sector had in common that the length of the professional development course is a barrier in completing the course objectives and be attentive to the courses they were teaching. An example of this was a participant trying to complete an online course, but he considered that the length and the number of units and modules were relatively extensive. Interestingly, the manager and a teacher from the public sector highlighted that the time spent in courses should be recognised by the institution or other authorities and this also demonstrated how teachers were interested in developing their teaching practices in the classroom. One participant also indicated that the ICELT was now being recognised by the Ministry of Education.

Lastly, several participants reflected on their own *English proficiency level*. It was also reported that participants' English language proficiency played an important role in gaining confidence whilst developing their assessment literacy throughout the assistance of professional development opportunities. Participants from the language institutes felt very comfortable with their English skills and communication, but some of these conducted the interview in Spanish. Once again, participants from the public sector highlighted that they were competent for the teaching courses, but that attending conferences where English was fully spoken could be difficult.

*'Sometimes those training courses are very extensive and they are instructed by English speakers and it is very difficult for me to understand them' ([INT] PUB-T3)*

The findings from this theme reported that EFL Tamaulipas teachers and programme managers across the sectors had different perceptions in relation to how they perceived professional development between academic backgrounds and teaching qualifications awards. In addition, participants demonstrated positive and negative attributions about the reasons why they were not interested due to different personal and professional aspects in language testing,

assessment skills, and practices. The following section reports participants and managers' perceptions towards their previous and current language assessment experiences.

## 5.4 Teachers' Previous and Current Assessment Experiences

This section presents the findings from participants' responses in relation to their previous and current assessment experiences. Participants mentioned they were able to identify and recognise which assessment methods, not excluding participants' teaching and learning practices, tended to work the best for their students based on their positive and negative experiences. Participants pointed out that they had used particular assessment methods and testing strategies, relating to the ones they found useful when they were language learners. All participants mentioned they were able to recognise if their tests were suitable for their language learners based on their previous assessment experiences. In other words, these participants believe that if something works for them, might work well for their students. This emerged feature (familiarisation) was also found among the three educational sectors due to the fact that they commenced their language learning journeys in a similar or the same institution where they were currently working. However, this familiarity with the current assessment practices might lead to a bias before, during and at the end of the assessment processes conducted in their classroom. This could generate influential decisions of selecting their current assessments practices. These participants might not have updated their assessing practices and the assessment and testing cultures have not developed according to the current needs.

The language programme managers from the public and language institute sectors also mentioned that teachers were responsible to recognise how assessment could be conducted based on their academic and professional experiences including how to assess their students' progress. This can suggest that participants should know how to assess their students appropriately, but not with the current trends.

*'I was a student in this school and the methods of assessment have changed (adapted), but not to a big extent... I used the same assessment practices in the classroom and for the teachers ([INT] LI-LPM)*

*'The methods of assessments are the same, the tests and some activities in class' ([INT] PUB-T2)*

Another emerged aspect was participants' assessment experiences on how they were assessed, and which methods worked the best for their language learning journeys. Participants mentioned that the way in which they received their feedback (relevant or irrelevant) and how they achieved their learning language courses had had an impact on how they currently deliver feedback to their students. Similar thoughts were presented in the public sector where participants also considered that their previous experiences were significant because they can

identify and form appropriate and constructive feedback. This was because they became aware and conscious whether they would need to develop their tests, other alternative assessments based on their institutional flexibility (operational assessment system) and the ability to design these tools. For private and language institutes, the assessment procedures demonstrated to be restricted or partial. Participants from the language institutes mentioned that they developed other materials for formative assessments, not only tests, but alternative assessments to facilitate the learning processes.

Participants were asked about their good and bad experiences during conducting assessments in the language classroom. For the public sector, they referred to good experience as the way they had developed their tests and materials because they underpin the contents for their assessments.

*'I believe that good experiences can enhance teachers to use certain tools... If I find it difficult for me, I might not use this...'* ([INT] PUB-T2)

*'A bad experience for me was the way teachers gave me feedback, I believe that this is so important and is a way in which learners can maximise their language proficiency'* ([INT] LI-T4)

These previous quotations demonstrated that participants were aware of developing their own instruments and selecting their contents, so their students could have obtained better results. This also referred to participants' perceptions towards what language learners required to successfully achieve the course contents and how teachers developed their instruments based on their previous knowledge and experiences. For some participants, especially from the public and language institutes, bad experiences were related to the way of delivering feedback to their students, avoiding information related to the structure and administration of tests (communication of assessments), the content and purpose of the tests (assessment purposes), avoiding use of testing resources due to participants' lack of confidence with other type of assessments (assessment processes) and the language ability for communication and evaluation (knowledge of applied linguistics and language). For private and language institute sectors, they indicated that good experiences for assessing their students were mainly by providing feedback to individual students (tutorials) and bad experiences were that they lack of access to be part of the selection of choosing the contents for test development and a large number of evidences and they were unable to decide which one could be essential for making relevant decisions. These thoughts from the participants could demonstrate that what they are currently teaching might not be similar from the content included in their tests.

Data reported that participants' experiences could play an important role on their currently assessment practices conducted in their language classrooms. The experiences as language

learners also could lead present teachers to find secure opportunities to assess their students but could also limit them to try other effective alternative sources for language assessments. Other relevant findings were found in relation to the lack of confidence in participants' English language proficiency, especially from the public sector. This is also discussed in the following section with the findings from participants' voices in relation to what they should know and do with language testing and assessment in their classrooms.

## 5.5 Teachers' Assessment Knowledge and Technical Considerations

This sub-section explores the knowledge and other technical considerations that teachers should know and do whilst conducting language tests. Three themes were emerged from the data 1) the quality and evaluation of testing materials, 2) the use of marking criteria for productive skills, and 3) other technical aspects like error correction, as well as the appropriate delivering of feedback and communication of assessments and results.

There were three participants from the public sector who mentioned they were responsible to develop their language tests and other assessment instruments. They stated that they occasionally conducted a validation activity to analyse and identify the quality of their tests. This type of validation process was intended to perceive whether a test could be useful for their students by conducting 'face validity' through sharing these instruments with other colleagues to receive some feedback and recommendations to improve certain aspects. PUB-T2 stated that *'asking teachers for advice'* could help other teachers standardise their assessment instruments in a systematic way for their institution. Also, this cohort claimed that there was no one who may possibly observes nor evaluates their own tests at their workplace such as programme managers or other local authorities. Participant (PUB-T3) shared her point of view that she believed some teachers did not take assessment seriously and most of the times teachers did not communicate or share with other colleagues due to a lack of confidence among teachers from the same institution.

The language programme manager from the public sector assumed that developing language tests was a challenging process, which teachers needed understand the different methods, types and item features they considered essential to be included in these instruments. Also, this manager stated that it was a teachers' responsibility to develop materials, mark and submit the students' scores and to communicate the scores to the institutional authorities and students. The other two language programme managers (the private and the language institute sectors) were responsible to develop and measure the quality of their tests. Both pointed out that they did not usually conduct any statistical method to evaluate their tests, but they were aware of techniques such as asking teachers for advice, or as the manager from the language institute who

commented they regularly '*pilot their tests*' with other teachers. Participants from language institutes shared the same previous point of view by informing that they were not included in the process of test development at any stage. Nevertheless, one participant mentioned that he has informed his manager if something was wrong with their language tests.

This can demonstrate that participants' perceptions towards the evaluation of their tests might not be similar among the three sectors. The flexibility in developing language tests from the public sector might create some bias at the moment of conducting these instruments, by the same token, the restriction from the private and language institutes might allow teachers to teach solely those language aspects considered for the test.

Participants were asked about the test usefulness and testing principles as part of their assessments knowledge in evaluating the efficiency of their language tests and the use of items. A total of 10 participants (all programme managers, the participants from the language institutes, as well as 2 from the public and 1 from the private sectors) were conscious of the language testing principles, with the rest mentioning that they were unaware of this terminology (n=5). This knowledge was acquired based on their academic trajectories and other courses they have attended as part of their professional development opportunities. However, a lack of knowledge about the usefulness of tests was found mainly from the three participants from the private sector and two from the public. The programme manager from the private sector highlighted that essential characteristics of tests were their authenticity and practicality because of the administration of standardised tests. Nevertheless, the two managers (the private and public sector) agreed that language tests should be practical and objective, referring to validity and reliability. Only the manager from the language institutes showed to have an extended knowledge about these five principles and construction of tests, and that he has applied appropriate the knowledge for developing test at the institution where he was currently working. Data showed that participants from the public sector had applied the principle of validity, more specifically 'face validity', to evaluate their language tests.

Since one important aspect of teachers' language assessment literacy is to develop high quality testing instruments to assess their students, skills and practices for tests and items evaluation, I asked participants for their experiences in item analysis. One participant from the public sector (PUB-T3) and the language programme manager from language institute demonstrated they were aware of the use of statistics to identify and improve inaccuracies with the language tests, but they mentioned they had never used this method before. In addition, these two participants were asked whether they would consider the use of statistics to improve the quality of the tests and other assessments. PUB-T1 and PUB-LPM pointed out that conducting

statistical analysis could be useful, but they would not be able to achieve this task since they were not skilled in tests analysis and there was a lack of time for conducting this activity.

*“I think by doing this practice [item analysis], it could improve the tests I develop but also other resources” ([INT] PUB-T1)*

*“The analysis of tests would bring lots of benefits for the institution and teachers, but teachers and I don’t really have enough time and possibly knowledge to do it” ([INT] PUB-LPM)*

The private and language institute sectors pointed out that conducting test analysis could be vital and it ought to be responsibility solely for the programme managers in those institutions, where tests were not developed by teachers. Two participants from the private sector and three from language institute pointed out they would like to conduct item analysis or evaluation of testing resources for their classroom activities they develop, but they were not knowledgeable for doing this task. The manager from the language institute had the belief that the use of statistics could only be used for those standardised tests since they have a significant impact on students’ progress, not for those tests administered in the classrooms. This indicates that item analysis might not be essential for language teachers but could assist them to maximise students’ performances and teacher’ skills and practices.

The next emerged theme was in relation to the use of criteria for productive skills. The four participants from the public sector pointed out that they did not use a specific criteria to assess these skills because they were unable to administer speaking or writing tasks due to time and the number of students in the classroom. However, PUB-T1 pointed out that sometimes students have class presentations, and these were marked based on the general performance, not always linguistically. Differently, two participants from the private sectors and all language institutes, highlighted the use of at least one or two criteria (holistic and analytic) that were given by their managers to assess students’ performances as a formal standard of their institutions. Two participants (LI-T1 and LI-T4) indicated that they have used of criteria sheets that were aligned to the CEFRL. PUB-T2 claimed that criteria sheets might help teachers mark students’ performances and products effectively based on different linguistic components such as fluency, pronunciation, organisation of ideas, content and coherence. Also, this participant expressed that it might be difficult to assess students if teachers lack knowledge on how to use rubrics.

An interesting commentary from the manager of the language institute was that they have a specific marking criterion that assesses students’ conversational skills. On the contrary, three participants (2 private and 1 public) mentioned that they considered that the use of criteria was not relevant because they knew how students might perform at specific learning level. They also pointed out that they tended to use a criterion they had practiced before. This could lead to

participants previous experiences in relation to choose what they believe is appropriate for their students. Two out of the four participants from language institutes pointed out that they split students' mark into the content and language skills, whereas the other two split students mark in different linguistic aspects (e.g., fluency, coherence, pronunciation, grammar and vocabulary) especially for students' written performances. The public sector cluster pointed out they do not often assess their students based on how they performed in relation to the contents or confidence, rather through language skills.

Another theme identified in all the participants was the assessment skills and practices of different error correction techniques, use of formative, summative and diagnostic assessments, and lack of confidence from teachers with their language proficiency. All participants indicated that error correction techniques were an essential factor that teachers should take into consideration for their teaching practice. The public sector indicated that error correction techniques were very useful for productive skills, especially for visual learners. The programme managers from the private and language institute shared similar thoughts of the powerful ways to allow learners to identify their mistakes in a friendly manner. A participant from the private sector, also stated that was relevant for their students' language proficiency, that by highlighting and signposting students' mistakes, they could learn better to use certain structures. One participant from language institute pointed out that for basic proficiency levels, error correction should be important, but also quite intimidated for students who are in higher levels of proficiency. By the same token, the private sector pointed out that techniques to correct students' mistakes relied on the students' proficiency level and teachers' techniques.

*"These techniques can help students to improve their English, I sometimes use finger correction; so students can see that there is something missing or not needed in a sentence..." ([INT] PUB-T1)*

*"I love marking my students correction but everything depends on the level, basic students is important since they are forming new sentences, but for higher levels cannot be that inspiring and this can lack their motivation" ([INT] LI-T3)*

All participants showed awareness among the differences for the diagnostic, formative and summative assessments. Participants from the language institutes indicated that these are administered in their language schools at different stages. Participants from the public and private sector opined that based on their language proficiency some teachers administer diagnostic tests, but these were not that frequent. Also, the private sector relied more on the summative assessments rather than the formative. A participant from this sector mentioned that the formative assessments were more the activities, exercises and practical tasks to complete. All participants including the language programme managers considered that the formative assessments were not that appropriate for students' progress since all these have summative

properties. In other words, participants expressed that even the formative tasks are counted to the students' final mark. A lack of formative assessment perception was discovered among the three sectors.

*"I am not sure if I use formative assessment adequately because at the end these ones also count to the final mark" ([INT] PUB-T4)*

Findings also highlighted that participants' current English proficiency levels could play an important role during conducting, marking and communicating language tests or other assessments. Three participants (2 language institute and 1 private) said that teachers might lack of skills to identify students' mistakes on time, and this could lead to an unsatisfactory use of the administration of language tests and other assessments. Two participants, (PUB-T3 and PUB-T4) confessed that due to their low English proficiency level, they had faced difficulties when delivering feedback, use of error correction or identifying students' proficiency in English. The language programme manager from the private sector indicated that some teachers should not teach higher levels if their language proficiency was not high enough, and it was under institutions' responsibility to hire teachers with adequate linguistic and teaching competences.

*'I am an English teacher but sometimes I feel afraid to write or speak since I do not feel confident with my level of English' ([INT] PUB-T3)*

*'All teachers are really good in here, but some of the language teachers did not have a professional or language education trajectory that can affect the way of correcting and marking their students' ([INT] PRIV-LPM)*

This sub-section discussed what participants should know about assessment, at the theoretical level, technicalities (e.g., testing principles, types of assessments, error correction) and evaluation of items were the most discussed. Also, different participants' points of views were highlighted at the moment of discussing the knowledge for the usefulness of tests and measurement theory. The public sector demonstrated to be less skilled than the private and language institute sectors due to their knowledge towards these technical considerations. In relation to participants' assessment techniques such as test development and the use of error correction techniques were the ones participants found important to know and do in the language classroom. Nevertheless, these findings also showed that teachers' language proficiency in English could lead to malpractices of assessments for the classroom. The following section reports on the findings from the participants' voices in relation to their assessment practices in the language classrooms, excluding language tests. In addition, it further explores the types of assessment, especially how teachers' perceive the formative and summative assessments.

## 5.6 Teachers' School-based Assessments

Participants were asked about what assessment practices, other than tests, had applied to assess their students' language performances. These findings were divided into two aspects: the assessments related to the class contents and students' progress (e.g., worksheets, rubrics, language portfolios), and the assessment for affective methods of learning (e.g., homework, attendance, participation). In relation to classroom assessment contents, all participants indicated that formative and summative assessments were both, equally important for the students' final mark, and that teachers used these two types of assessments to evaluate students' language learning journey. Data showed that participants used these terms (formative and summative) differently. Participants from the private sector and language institute, expressed that the term "formative" referred to those types of assessments that are not structured as tests; whereas, "summative" mainly deals with all those sources, including tests, that are linked to the final evaluation of the students. For the public sector, the formative concept was the stage in which teachers provide feedback to students, solely to improve the quality of students' progress, and prepare them towards the final test. Two participants from the public sector (PRIV-T2 and -T4) indicated that they did not use other sources apart from language tests or mock tests to assess their students. On the contrary, the rest of the participants (n=13) mentioned that they used *language portfolios* which mainly included learning tasks and formative assessments as part of their students' evaluation. Data also showed that three participants from language institutes mandatorily require to assess students through language learning portfolios as part of the institutional criteria.

Some other assessment sources used by the participants in their classrooms were *worksheets* as an example, PUB-T1 stated '*I develop several worksheets with the contents from the class before, I think this is a good way to test students whether or not know they learnt what we did previously*'. Similarly, PRIV-T2 and PRIV-LPM pointed out that they develop certain worksheets, but they were closely as '*quizzes*' and were mainly for summative purposes. Another method for assessing students' performances was throughout '*individual or group presentations*'. Three participants for language institutes used presentations for the purpose of assessing speaking skills, for both formative and summative.

All participants among the three sectors indicated that they selected certain exercises from students' course books and/or workbooks. Only one participant indicated that the institution did not use a book and was required to develop and adapt materials from different resources for the students such as internet resources and relevant books. The rest of the participants mentioned that they used workbooks for formative purposes, but they indicated that they had selected certain tasks with the intention of including it as summative. Participants (PRIV-T2 and PRIV-T4)

also mentioned that the use of *'journals'* and *'essays'*, which participants considered vital for assessing students' writings, were included as summative assessments. In addition, three out of the four participants from language institutes pointed out that they had used methods such as the use of an *'online platform'*, in which students have to complete several activities and that the participants had marked and provided some constructive feedback. This online platform is the one that includes the course-book. The language programme manager from the language institute claimed that online platforms were useful, but not all students could have easy access to these, due to the lack of technical equipment. The language institutes mentioned that they developed materials for formative assessments, not only tests, but alternative assessments to facilitate the students' learning processes.

*'Most of the formative evaluation include a selection of activities from the students' course book and sometimes some worksheets I found on different websites' ([INT] PUB-T2)*

*'I ask my students especially from intermediate levels to write their routine, simple things they can easily write and learn daily vocabulary' ([INT] PRIV-T4)*

*'I usually upload some materials [to an online platform] for testing my students' knowledge and I can provide some feedback once these activities have been finalised ([INT] LI-T1)*

There were other aspects that influence students' scores, as well as other products. The use of homework and classroom participation were factors that these participants took into consideration for summative purposes. Two participants from the public sector mentioned they sometimes take into consideration external factors such as students' motivation, the effort that they had made and their progress. For language institutes, two participants mentioned that students had to take some extra work, via a self-access centre, at their institutions, but that this was not considered towards their final mark (summative assessment).

A participant from the private sector pointed out that students' attendance to the class, conversation club and independent work were part of the final mark. Similarly, participants from the private sector mentioned they register students' attendance as part of summative assessment. A participant from the private sector also believed that the assessment practices might vary from one teacher to another. The accessibility regarding the use and development of tests, preferable methods for assessing, the criteria participants used and alternatives in assessments could enhance teachers understanding on how to manage assessments to improve and boost students' learning trajectories. The following section reports on the findings from participants critical issues in assessments per educational sectors that they have perceived.

## 5.7 Critical Issues among Educational Sectors

This section has been divided into three sub-sections, each one reporting participants' educational sectors in relation to the assessment critical issues or problems they have faced in their language classrooms.

### 5.7.1 The Public Sector

This cohort indicated four challenging situations during assessing their students. These were 1) the number of students in the classroom, 2) students' progress in the institution, 3) lack of support from their managers or other authorities, and 4) external assessment features. The four participants pointed out that assessment was difficult to conduct since teachers had a significant number of students in the classroom, as well as several groups depending the position they hold: part-time or full-time. One participant mentioned that it was difficult to assess students because of the lack of time they have in the classroom. Since assessments were an important aspect for students' learning, not all teachers could conduct alternative assessments and provide feedback to individual students. Likewise, there was a lack of productive skills which were not frequently tested because of the number of students in a classroom and the time it demands to mark these.

*'it is so difficult for me to provide feedback when I have to deal with 50 students in just 90 minutes.... I prefer to develop tests so students can easily distinguish what they need to learn' ([INT] PUB-T2)*

*'...some of my students have good level of English, but some of them struggle so much in different aspects they supposed they had learnt in previous courses. The previous teacher might have not evaluated them well' ([INT] PUB-T1)*

The manager from this sector pointed out that students' progress, from one academic term to the next one, was highlighted mainly as a problematic and difficult to track. Three participants indicated that other colleagues were unable to assess their students accordingly to their syllabus and that could affect their students' progress. One participant suggested that teachers should coordinate with other teachers to analyse students' current language levels in order for teachers to share their thoughts and experiences. Participants also indicated that they would like to have more guidance and support from their managers and other authorities in terms of preparing teachers with updated materials and methods to assess and teach their students. At first point, two participants indicated that it would be beneficial if managers could bring the tests rather than being developed by the teachers. On the contrary, the other participants mentioned that it would be essential to revise testing instrument as a group of teachers. They did not demonstrate any dislike to the process of developing tests, but this could be a factor to empower students' progress based on the institutional policies.

Lastly, two participants stressed their concerns in relation to external assessment factors such as; behaviour, attendance and homework responsibility as part of their assessment components. These two participants considered that these features of assessments should not take into consideration, but the institution demanded these as essential factors, since these are vital to form new professionals, not only in the field of language education, but in general aspects of education.

### 5.7.2 The Private Sector

Similarly, as the public sector, this cohort expressed their concerns for students' progressions that should be tracked and measured adequately. Participants indicated that students should not be able to register to any of the courses if they lack the English proficiency level. Two participants pointed out that students should not be accepted until they have the English requirements for admission and graduation purposes. This sector made emphases on students who had just finished their upper-secondary school and jumped to tertiary education level. According to participants' students' experiences, they had struggled to obtain higher scores since their testing instruments were developed based on international standards, but not based on how the teacher has taught the class.

The other problem was that participants disliked the fact of not being considered as part process of developing and selection of testing materials. They suggested this could bring several benefits for their teaching practices since each educator teaches and assesses differently. This could lead to adequate testing materials for each individual group, but they also indicated that was also a great perspective for measuring the English language based on international standards which recognise the institution.

*'...students from high school (upper-secondary) are assessed differently as university students (higher education institution). This big gap shows how bad teachers were in terms of assessment to promote these weak students...'* ([INT] PRIV-T3.

*'...speaking tests should be taken into consideration, regardless the fact that summative tests do not include this skill...'* ([INT] PRIV-T1.

Moreover, two participants indicated that the lack of speaking activities was a problem since students read and write regularly, but the factor for speaking skills such as pronunciation and fluency were rarely assessed, and that not all teachers do this important activity. Also, a participant from a tertiary educational institution showed that speaking should be taken as formative since speaking was not included in the final examination. This could lead to empower English classes with better resources to teach and assess speaking skills accordingly to current needs and appropriate methods.

### 5.7.3 Language Institutes

The four teachers from language institutes pointed out that the use of language portfolios was a critical problem since teachers had different perceptions on how to use this tool. One participant mentioned that the use of portfolio should be discarded because this tool was not what the institution required. A participant (LI-T4) pointed out that *“the value of collecting learning evidences is more related to put everything they have done during the academic term”*. This was an example that highlights how teachers can use a tool differently and affects students' progress. The use of portfolios was highly used in language institutes because they follow the parameter of the CEFRL and this included a language portfolio. A participant claimed that training on how to use portfolio, as well as other sources of assessment, could improve their teaching practice. One participant also said that it is essential to understand the qualities of assessments, so as teachers should be more flexible within the assessment practices.

Other critical issue was the methods of assessing the productive skills, mainly writing. The use of holistic and analytic marking schemes should be taking into consideration. Participants highlighted that the lack of marking criteria could lead to a bias performance. A participant explained her concern about her lack of experience in assessing writing with the institutional marking criteria. She informed she never received a training from the institution in relation to use these marking tools.

*“Marking writing is the most complicated task in here [institution], I never received any training to use the criteria because it is complicated and I relied on my personal experience” ([INT] LI-T1)*

*“The use of portfolios is irrelevant, teachers just give students many pages to read and these are never looked again by the students... portfolios are useless” ([INT] LI-T3)*

Examining the critical issues in assessment that teachers have faced could highlight the educational sectors current misuse of assessments; and therefore, this could assist to empower students' progress at local, but also at national level. Participants from the public sector indicated concerns for external factors (e.g., the number of students in the classroom, external factors of assessments and the lack of support in developing testing materials) when assessing students' performances. The private sector and language institutes were more concerned on students' performances such as the use of language portfolios, the lack of some skills to be tested and the progression of the students with regards to previous language courses. The following section presents a summary of the Phase 2 of data analysis.

## 5.8 Summary of the Chapter

This Chapter reported the findings from the Phase 2 (semi-structured interviews). Semi-structured interviews were analysed through a thematic analysis in which there were seven

themes. This was to answer the research questions from the exploratory research (1, 2 and 3). It is worth mentioning that these findings were also triangulated with the TALDI findings. Some of the main findings were the following:

- The first interview theme, emerge from the data, was in relation to the tests, their importance and occurrences. EFL teachers in Tamaulipas administer language tests to observe how students have progressed in a period of time; this can be per month, every two months or at the end of the term. The number of tests conducted in language courses vary due to educational level (high-school or higher education institution). The number of tests also links to the purpose of the tests, to diagnose students' language proficiency level at the beginning, to measure students' progress as formative tests and to make educational judgments whether students have learnt enough to be promoted to the next learning stage or completion of a course. This theme also deals with the status quo of language assessment literacy that allows institutions to visualise the reasons of conducting assessments.
- Participants tended to assess different language skills and features based on their teaching practice. The public sector mainly assesses grammar and vocabulary, these participants think that the way of marking and provide feedback is much accurate and slightly effective. The private sector included most of the language skills and they manage to include integrated skills to assess linguistic components as grammar and vocabulary. Nevertheless, the private sector does not tend to assess speaking skills. The only sector that demonstrate an extensive method of evaluation was the language institutes in which these participants have the four language skills tests plus a complementary section that involves grammar and vocabulary. The development of the test is mainly from the language programme managers in the private and the language institute sectors. Similar findings were reported from TALDI in relation to test responsibilities and the contents.
- Participants indicated that professional development was a possibility to reflect, to work collaboratively with other teachers, and to adapt new teaching and learning opportunities into their language classrooms. However, they did not consider language assessment as a priority for their professionalism. There was a discrepancy between the private and language institutes based on the facts that CPD opportunities should not be mandatory. Tamaulipas EFL teachers also consider those teaching qualifications as part of their professional development or trajectories. These teaching qualifications also dealt with the hiring and recruitment processes, but these change based on the sectors. Participants stated that there was a lack of instruction of language assessment from their teaching qualifications and their academic programmes. This theme was relevant to answer research question 3, the backgrounds of professional development towards language assessment

literacy. As a consequence, participants did not frequently attend trainings towards improvements in assessment, but they prefer courses that include strategies and techniques to deliver knowledge (language)

- Participants' previous experiences such as identification of qualities to give constructive feedback can lead to teachers to accommodate elements to have a solid and congruent recommendations to empower students' performances. EFL teachers in Tamaulipas have expressed that the way in which they have been assessed has also influence the decisions and methods which they are currently assess their students.
- The knowledge and technical consideration of language assessment from participants seem to be significantly diverse. Several participants were unaware of the testing principles or usefulness of tests, but the ones who had some notions of this relevant topic knew how to apply in basic technical operations. None of the participants have used educational measurement theories to identify the level of difficulty, discrimination of their tests. Participants' awareness of the three types of assessments: diagnostic, formative and summative were notable among the three sectors. However, participants did not inform of other essential aspects regarding knowledge of assessment such as national and local assessment policies, knowledge of applied linguistics nor educational theories. Comparing to the findings from TALDI, knowledge of assessment was different from the public sector than the others, the private and language institutes.
- The use of formative and summative assessments inside the classroom varies significantly among the three sectors. These two types of assessments are sometimes considered essential whilst counting students' final mark. Some participants mainly from language institutes and the public take into consideration worksheets and language portfolios as part of their summative assessments. Two participants from the private sector also highlighted that they just provide feedback to students, but they do not track whether the student has improved based on the feedback given. Two types of assessments were considered classroom assessment and affective learning assessments.

Some of the critical issues mentioned by the participants from the public sector were the number of students in each classroom, the lack of support from the school authorities and the way in which participants assess students' affective factors such as behaviour, students' responsibilities and homework. For the private sector, the students' progress from the previous and current courses is not properly tracked and they were several problems with the curricula since it was not progressive. In addition, the fact that teachers were not considered for the purpose of test development and a lack of preparation of tests. Participants from the private sector seemed less prepared in language testing since they do not develop tests. Finally, for the language institute

## Chapter 5

sectors, the assessment of portfolios was not properly assessed by the teachers; they do not provide a continuing formative assessment to distinguish whether students have progressed based on the feedback provided by their teachers.

## Chapter 6 Workshop: Progress and Impact

This Chapter reports on the findings from the experimental research conducted via the instruction of a workshop (Phase 3) and measuring its impact through a focus group interview (Phase 4) with the intention to analyse and explore what areas are considered more and less important in relation to language assessment literacy among EFL teachers in Tamaulipas and to measure its impact on participants. The first section presents the quantitative findings from the pre- and post- measures that were reported based on the components of these measures, educational sectors and participants' progress. Section 6.2 links the previous findings with the relevant outcomes from the practical and reflective tasks. Section 6.3 presents the evaluation of the workshop which scrutinises different aspects of the workshop as the contents, materials and instructor. Section 6.4 reports on the impact of the workshop on participants with a qualitative analysis from a focus group interview. Lastly, this Chapter concludes with a summary of the relevant findings from the experimental research to be discussed in Chapter 7.

### 6.1 Pre- and Post- Measure

This section addresses the findings from the analysis of the two measures (pre-measure and post-measure) administered at the beginning and the end of the workshop. This measure was developed based on 9 language assessment components: a) concepts in assessments, b) testing principles, c) types of items and tests, d) development of testing instruments, e) marking schemes, f) item analysis, g) communication of assessments, h) alternative assessment, and i) assessment as educational support. A total of 40 items were chosen to identify participants' knowledge before and after the workshop. This measure can be found in Appendix H.

Table 32 illustrates the components of these measures which were reported with minimum and maximum values, the mean of the component (***M***) and the standard deviation (***SD***). This table also includes the difference captured, from the pre-measure and post-measure, in each of the components of the construct. In addition to these figures, the rank of difference is also reported in the last column of the table.

Participants showed that the component of *item analysis* obtained a difference of average 3: pre-measure (0.9) and post-measure (3.9). This component obtained the highest number in difference placing it at first aspect with the most difference captured. *Types of items and tests* obtained an average difference of 2.8: pre-measure (4.4) and post-measure (7.2). The following component *development of testing instruments* obtained an average difference of 1.7: pre-measure (3.9) and post-measure (5.5). The components of *testing principles* and *alternative assessment* obtained averages of 1.2 and 1.1 respectively. *Communication of assessments*

obtained a difference of 0.9. *Concepts of assessment* and *assessment as educational support* obtained a difference in average of 0.6 from both measures. There was no change in difference reported from the component of *marking schemes* since participants obtained these items correct in both measures, this might suggest that participants had already some notions to distinguish from holistic and analytic criteria to assess productive skills in the classroom.

**Table 32 Overall Findings from Pre- and Post- Measures (n=10)**

Components of the Measures	Pre- Measure				Post- Measure				D*	Rank
	Min	Max	M	SD	Min	Max	M	SD		
Concepts in assessment	2	3	2.4	0.52	3	3	3	0	+0.6	7
Testing principles	2	4	3.4	0.70	4	5	4.6	0.52	+1.2	4
Types of items and tests	3	6	4.4	0.97	6	8	7.2	0.63	+2.8	2
Development of Testing instruments	2	6	3.9	1.20	5	6	5.5	0.53	+1.7	3
Marking Schemes	3	3	3	0	3	3	3	0	0	8
Item analysis	0	2	0.9	0.74	3	4	3.9	0.32	+3	1
Communication of Assessments	1	3	2.1	0.57	3	3	3	0	+0.9	6
Alternative assessments	2	5	3.9	0.99	5	5	5	0	+1.1	5
Assessment as Educational Support	1	3	2.4	0.7	3	3	3	0	+0.6	7

\*Note: D stands for difference of means captured.

The following step was to identify discrepancies, if any, among the assessment components of these measures with participants' educational sectors. Table 33 reports the findings from participants' educational sectors (4 public, 3 private and 3 language institute) and each component with the minimum and maximum values of correct responses, the mean (*M*) and the standard deviation (*SD*). Findings from the participants' educational sectors and the components of the measure are reported in Appendix L.

**Table 33 Findings from Participants Educational Sector (n=10)**

Participants' Educational Sector	Pre- Measure				Post- Measure				D*
	Min	Max	M	SD.	Min	Max	M	SD	
Public Sector	21	28	24.5	3.51	36	38	37.25	0.96	+12.75
Private Sector	24	30	26	3.46	39	39	39	0.0	+13
Language Institutes	29	30	29.3	0.58	38	40	38.67	1.15	+9.34

\*Note: D stands for difference of means captured.

Participants from the public sector scored the lowest mean of 24.5 out of 40 at the pre-measure; conversely, the post-measure showed there was an increase of participants' progress by obtaining an average to 37.5 (+12.75). Data showed that participants from this sector were less skilful/knowledgeable in certain assessment components at the beginning of the instruction. Participants demonstrated an increase in progress in components of *item analysis* (+3), *types of items and tests* (+2.75), *development of testing instruments* (+2.5), *alternative assessments* (+1.75) and less progress was captured in components like *testing principles* (+0.75), *concepts of educational assessment* (+0.5), *assessment as educational support* (+0.25) and the *use of marking schemes* reported zero.

Participants from the private sector recorded an average of 26 at the pre-measure. This group of participants obtained a slightly higher difference than the public sector (+1.5) and they achieved a mean at the post-measure of 39, making this a progress capture of 13. Progress was highlighted in components such as *types of items and test*, as well as item analysis with an average of (+3.33), *educational assessment as support tool* (+1.67), *alternative assessments* and *testing principles* with (+ 1.33), but less progress was observed in *concepts of assessment* and *communication of results* with (0.67), and *marking schemes*, with similar scores as the participants from public sector with zero.

In relation to participants from the language institutes, a higher average of correct items was obtained at the pre- measure with 29.33, and at the post- measure showed an increase to 38.67. It is important to mentioned that one participant from this cohort obtained the maximum number of correct responses in the post-measure. This group reported a difference between the both measures of (+9.34) resulting to be the group with the lowest discrepancy. However, participants showed progress in components like *item analysis*, similarly as with the public and private sectors, showing a difference of (+2.67), *types of items and tests* (+2.33), *testing principles* (+1.67) and *development of tests* with (+1.00), but no difference was reported on *marking schemes* and *alternative assessments* with zero between the measures.

These findings demonstrated that participants who obtained higher differences between the scores from both measures, to some extent, made progress, but this needs to be compared to other aspects such as the practical and reflective tasks to identify which areas participants made significant progress. However, those progressions pathways that were observed with lower difference does not necessarily reflect that participants did not make any progress in understanding the components of the measure. In fact, it gives the assumption that those participants were conscious of such elements before the instruction of the workshop and empower their skills with the assistance of the workshop. In addition, at individual level, all participants showed improvement, this can be found in Appendix N. These findings were compared to the ones from the findings of participants' reflective and practical tasks. The next section presents the findings of participants' tasks and that these have been reported on participants' educational sectors.

## 6.2 Reflective and Practical Tasks

This section reports on how participants performed during the development of the reflective and practical tasks. In addition, this section informs what participants did in the practical tasks and what they considered during their reflections. Participants were required to write three reflections and comment on how they performed on the two practical tasks. These

reflections were analysed through thematic analysis based on their participants' educational sectors. The full description of the reflective and practical tasks can be found in the syllabus Appendix G and an example of the reflective task written by one participant can be found in Appendix I.

### 6.2.1 The Public Sector

In the first reflective task, participants had to write a reflection about the importance and methods for language assessment in their classrooms. This group indicated that assessment was more towards the process of marking students' performances and how much their students have learnt in a period (e.g., monthly, every two months or per semester). Importantly, three participants' reflections showed they used assessment in most of their classes by using different resources, not necessarily tests, but a diverse number of alternative assessments such as portfolios, worksheets, error correction techniques and selected activities from students' workbooks. In addition, participants stated that there were two other aspects that were considered during assessing their students: participation in class and assigned homework. Participants also highlighted difficulties since they tend not to have accurate methods to properly assess these external factors.

Two participants also declared that students' motivation sometimes was considered a factor to take into their final mark, to the extent of upgrading students with poor linguistic ability. This clearly indicates that teachers might upgrade their students even if their language skills are not appropriate. Participants considered that the best methods to assess language learners were the ones which students feel much comfortable such as group activities or presentations in which students could interact and work collaboratively with others in the classroom. An example of a reflection from one participant from the public sector is below. This participant is aware that assessment can bring benefits to highlight and to spot students who have made progress and those who have not.

*'...assessment is the way to see if my students learnt or not what I taught during the class. It is very important because it helps us identify students who need more help and preparation. I usually have (language) tests, worksheets, and pages from the workbook, attendance and participation. The test is 30 % of each partial [two-month period] and 70% the other. Students like communicative tasks but it is difficult sometimes, I have 40 students...'* ([REF1] PUB-T3)

The second reflective task was related to the development of tests and their efficiency. This reflection was in relation to the first practical task. Participants reflected on three fundamental aspects to take into consideration at the moment of developing tests and alternative assessments: a) the level they are teaching, b) the contents that were taught in the class, and c) learners' preferred methods for assessing. Although one participant indicated that only language

proficiency was the most important element, as well as its usefulness towards the students; this participant was more oriented towards the level of the class. After the development of the first task, this group found the use of item analysis was challenging, but useful to identify the efficiency of different tasks, not only tests, but also other alternative assessments such as tasks, worksheets, etc. They suggested all educational contexts should evaluate their testing resources to identify whether the methods of assessments were appropriate for their institutions, but two of these participants pointed out that this could consume a lot of their working time. The other two participants agreed that they would conduct item analysis to facilitate the future development of teaching resources and to develop a series of useful tests for future courses. Participants also indicated that item analysis could bring benefits to students' performance and progress, as well as to evaluate and reflect on their own teaching practice. The level of reflection from this participant can demonstrate that the workshop was useful to make awareness that teachers can also be privileged to identify any gaps in their teaching practice. One participant stressed the importance for item analysis for educational purposes to identify the alignment of the contents and the type of items and tasks with the intention to identify whether the methods of teaching were according to what students were tested.

*'...evaluating the items of the test we created was very challenging because I did not know anything about statistics or even calculations to identify if items are well developed for the tests. This is very useful and I think I could get a lot of information to improve my tests and ways of how I teach...'* ([REF2] PUB-T1)

The last reflective task was in relation to problematic situations in which participants have dealt with when they conducted assessments in their classrooms. The public sector pointed out three main problems: 1) the lack of English proficiency level (students who did not belong in the same language learning stage as others), 2) the number of students in a classrooms, and 3) the misuse of assessments to improve students' learning. Participants were also able to reflect about possible solutions to their own problems conducting language assessments in their classroom. A participant (PUB-T4) suggested she would administer diagnostic tests at the beginning to capture students' proficiency levels. This participant also mentioned the lack of confidence of her own language proficiency in English. PUB-T5 could examine their students before the class to evaluate their knowledge and to possibly adapt better teaching, learning and assessing resources to her class. For the number of students in the classroom, a participant suggested that it was vital to divide the group from advanced to less advanced students; so the advanced students could also support their classmates. Finally, this participant also stated that the misused of tests and assessments could be avoided if teachers were well-trained and prepared with relevant testing resources.

*'...the English level is the main problem for me and my students, I am teaching intermediate level where I develop my tests and most of my students seem to fail. I think the tests I choose are very difficult. Sometimes I think my English is not good enough to be able to identify my students' mistakes, especially in writing...'* ([REF3] PUB-T4)

In terms of the last practical task, one participant (PUB-T2) conducted an evaluation of a writing task which had to evaluate two students' pieces of writing using a criterion based on the B1 (CEFLR). This participant noticed the differences in marking before and during the workshop. The participant considered this workshop as an opportunity to improve the way to mark and provide relevant feedback to their students' performances. Another participant (PUB-T4) evaluated the methods for feedback delivery, especially for speaking skill. This participant brought two different recordings from her students and considered that the feedback given before to the student did not provide a constructive reason for improving. In fact, this participant also mentioned that the previous methods for feedback were to spot mistakes, principally grammar and pronunciation. This participant also considered feedback as a crucial feature of assessment to improve students' language progressions.

The remaining two participants evaluated the usefulness of worksheets in the classrooms. They pointed out that a worksheet should be developed based on the task specifications that actually matches the learning course objectives and the students' preferences for learning and practicing. They both agreed that worksheets could not be measured as part of summative purposes, due to the lack of learning evidence. The lack of elements to make judgments on how participants progress and advanced during the course with alternative assessments could not benefit the learning progress. The four participants demonstrated that they progressed, improved skills and practiced in areas of test development, evaluation of items, and the use of alternative assessments. They mentioned that assessment was important, and it was used almost daily in different forms.

*'...the feedback I give my students is on how to correct their mistakes. After the session of receptive skills, I had the need to help my students to learn better than learn by heart. I could give her (student) better feedback and advice to improve certain structures while speaking...'* ([FGI] PUB-T4)

These findings also concurred with the ones in which participants from this cohort improved their knowledge from the pre- and post-measures. Participants from the public sector expressed their awareness and the importance of knowledge of assessment for better skills and practices to boost language teaching and learning practices. The following section presents the findings from the analysis from the private sector.

### 6.2.2 The Private Sector

Participants from the private sector pointed out that assessments in their institutions were dynamic elements for their teaching practice since language tests were considered potential tools which decided whether students were able to advance to higher levels. One participant indicated that the assessment practices were not as frequent in their classes, but they were essential factors for teachers to make educational decisions for exit requirements for undergraduate programmes and selection of new students for undergraduate programmes.

One participant (PRIV-T3) pointed out that the scores that teachers provide to students reflect could also reflect on their teaching practice. PRIV-T3 reflected on the way that the students' scores have a serious relation to the way teachers perform in the language classroom; in other words, the better teachers teach, students have a potential opportunity to get an appropriate score. All participants highlighted that language tests were the main tools for assessing their students, with only one indicated the use of other alternative assessment to maximise their students' learning journey such as presentations, homework and group work.

The reflections showed that private teachers tended to rely purely on test-scores mainly from standardise tests such as TOEFL. They indicated that they would like to incorporate other resources for measuring students' performances (e.g., workbooks, presentations, portfolios), but these were not currently recognised as summative assessments in their institutions.

*'...test are very important since students take them to fail or pass the classes. I used language tests that were developed by my coordinator and students have to pass them. I used other resources such as quizzes and sometimes I ask my students to develop a presentation in teams... ([REF1] PRIV-T3)*

Several participants reflected that the practical task one was not relevant since they did not develop their tests, but some other alternative assessments such as quizzes and presentations could potentially empower these alternatives by an adequate item analysis. All participants indicated that the evaluation of items and tests were useful tasks for those teachers who were responsible in developing their own tests and for language programme managers. All participants highlighted the importance of test development so as to inform their school authorities about their students: how students learn, and which task could be more suitable for their learning. Also, one participant (PRIV-T2) pointed out that some of them also worked in other institution while attending the workshop, where test development was required by the teachers, and therefore, this participant highlighted that the evaluation of tests could lead to time consuming, but vital for the development of fruitful testing resources according to one specific group. Although, private participants found the use of statistics useful for educational purposes, they all discarded the possibility to conduct item analysis unless they were required to do so.

*'...this task was difficult but I enjoyed it, my last course about numbers was in high-school. I never thought I would use number for teaching English. I am working in two schools XXXXX and XXXXX, where I help developing tests. I am thinking of using this in my free time, but I don't think I will use it for every test, it would be endless work...'* ([REF2] PTIV-T2)

In the last reflective task, participants highlighted their disappointment of not being part of the process for developing language tests for summative purposes in their institution. A participant (PRIV-T1) highlighted that some learners should not be in some higher levels since they were unable to verbally communicate and develop writing tasks appropriately. Some of the solutions participants included in their reflections were that teachers could provide an individual report of students who required further improvement and inform the language programme managers about how their students have been progressing during the class in order to increase awareness to the authorities about their current learning situations and spot any students at risk of failing. Also, it was suggested that teachers should be more flexible during assessing students to distribute the final mark within other assessment methods, especially alternative assessments where they can track their own students' progress.

*'...some of my students, especially in higher levels, do not have the essential language aspects to communicate. Their grammar and vocabulary are poor and I believe these students should be reported to the coordinators. These students are not learning and it is sometimes difficult for me to adjust my class for different levels...'* ([REF3] PRIV-T1)

For the second practical task, two participants evaluated the process of feedback to be more flexible and constructive mainly on students' homework to maximise the formative assessment. It is important to remember that the participants from the private sector have less control over students' assessments. A participant evaluated a piece of her students' homework to identify methods to provide better and accurate feedback and to examine what students require for the English course they were currently undertaking. This participant considered that homework could be an essential factor to track students' progress and to explore what they need to obtain better results in the language tests, summative assessments. Another participant evaluated the purpose of learning portfolio to explore the positive and negative aspects towards students' progress and learning. This participant reflected on the error that sometimes teachers do not see language portfolios as an opportunity to check students' progress, in fact, this participant stated that the use of portfolio for the classroom was more as a device for collecting different learning activities. Due to the lack of testing resources in the private sector, these three teachers evaluated the way in which they could incorporate alternative solutions to improve the formative purposes of materials for their students. Participants felt satisfied to use alternative assessments, but they also pointed out that this could be a lot of work, not only for the students but teachers as well. Although teachers should be able to develop individual reports for their

students' performances based on the use of language portfolios, they indicated that it might take too much time from their teaching and planning time.

Participants from this cohort demonstrated they increased knowledge, based on the findings from the measure in technical abilities such as test development and item analysis, but they also showed their discomfort for not being considered for the purpose of test development in their institution. Tensions between these language teachers and their school authorities were discovered in this group. Nevertheless, they also showed improvement in alternative assessments such as portfolio, homework and presentations to improve their students' proficiency and learning methods.

### **6.2.3 The Language Institute Sector**

Language institute teachers highlighted that assessment was an important factor for making institutional decisions such as upgrading their students to the next learning stage and identifying what learners have acquired. Assessment seems to play an important role in this sector, and teachers tend to be the ones who are partially involve in this practice. Participants mentioned that they have control from the 50% to 60% of the students' progress, which is reflected as a formative assessment. Also, participants were the only ones who mentioned that autonomy was part of the process of learning in these institutions, since some participants had a special place where students can learn and improve their skills by taking some extra hours in a self-access centre.

A participant (LI-T1) included that assessment was also part of their performance; students in this participant's institution tend to assess the teachers work at the end of each course. It seems that assessment in this cohort has more possibilities to unify assessment practices because of students', teachers' and language programme managers' needs. Although participants indicated that language tests were developed by their school authorities, language teachers were in charge in developing several alternative assessments such as worksheets and journals, materials to be used under the learning portfolios and rubrics to assess students' academic products. Each of these alternative assessments were utilised towards students' formative assessment, but also these previous ones weighted several points for students' final mark. In order words, the tests (developed by their managers) plus the alternative assessments were combined to obtain the students' final marks. Besides this, two participants indicated that the process of selecting which testing and assessing resources work better for their students relied on their experiences as learners. These two participants were the only ones who mentioned experiences as part of their selection assessment methods.

*'...assessment is very important not only for students but for us as well. As teachers we are also assessed by students on how we performed as teachers. In terms of language tests, I do not develop them because my boss makes them for all levels. These are given to us the day of the test. We are responsible for promoting learning with portfolios that count 50% of the students mark...'* ([REF1] LI-T1)

For the second reflection, participants pointed out that they take into consideration two main factors for developing testing resources, the level of the course they are teaching, and the previous feedback delivered to students. These two play a crucial role because they tend to develop new sources for each group and the level. Hence, their alternative assessments tended to be oriented to what students required based on each specific language learning level. At the moment of conducting item analysis during the workshop, all participants indicated that this could bring a lot of benefits, not only to identify whether a test was useful, but to guide teachers in the integration of better resources for their assessing practice. However, these participants had no access to the development of the tests (similar to the private sector), they constantly evaluate the test contents and these were reported to the authorities if these are not accurate. Only one participant mentioned interest towards the use of item analysis for improving education practice, especially for teaching students how to choose the correct answer in certain items. Participants indicated that the evaluation of items (item analysis) could be a responsibility for language programme managers to bring better sources to the institution, both for students and teachers.

*"...this was epic, difficult but relevant to understand the process of test development. I used to create items and tasks from my own personal experience, contents and the level. Never expected to analyse a test in the way we did in the workshop. It is very useful but this must be a task for my coordinator who develops the tests..."* ([REF2] LI-T3)

In the last reflection, participants express two main problems while conducting assessment in their institutions. The misuse of several alternative assessments and the unfair process for marking productive skills. Some of the solutions for the large number of assessments was to reduce the number of tasks by gathering teachers to work in academies or groups, in other words, those teachers who teach the same level could select a repertory of activities that could assist students to obtain the learning objective adequately. In addition, one participant pointed out that the use of portfolio should only contain elements that were related to the learning course instead of a portfolio of all learning evidences. One participant pointed out that students and teachers sometime misunderstand the purpose of a learning portfolio. Another solution for this was to keep individual record with a particular student's product (activity to be further revised). With regards to improving teachers' methods for assessing productive skills, the implementation of developing accurate rubrics for each task could guide students to achieve the purpose of the tasks in a contextualised aspect. One comment of this participant (LI-T2) was that *'...I have my own*

*experience assessing speaking. Never been guided if what I am doing is right or not*'. This demonstrate that there was no training for assessment in this institution.

For the last practical task, this cohort evaluated and examine several aspects related to the assessment in the classroom. One evaluated the efficiency of using journals for improving students writing. Although it was difficult to assess and provide feedback in each of these sources, this participant found interesting the fact that students could practice and expand their lexicon. Other (LI-T3) examined the use of workbook activities as summative purposes which were not reliable for summative purposes since these demonstrate how much students have gained in a particular lesson. Also pointed out that these workbook activities for summative purposes would not match the expectations of the course, nor the curriculum. Finally, one participant examined how the conversation club in the institution could improve students' speaking practice, as well as its assessment. This participant pointed out that this could be a great perspective to track and assess students because they could gain different vocabulary and be ready for a proper speaking examination, but the attendance to this workshop should not indicate that the students have improved. Participants from this cohort showed interest in aspects of item analysis and alternative assessments. Their interest to understand how tests could contribute towards the process of learning was mainly notorious. The following section presents the findings from the workshop evaluation sheet that includes the findings related to the objectives of the course, contents, tasks and materials, and the contextualised component of assessment and evaluation in the Mexican State of Tamaulipas.

### **6.3 Workshop Evaluation Sheet and Focus Group Interview**

The purpose of having this evidence was to examine and explore the efficiency of the workshop and the materials developed and produced before and during this intervention since they could potentially provide vital features to discover how teachers perceived and performed during, but not only this one, training course. The quantitative findings from the participants' workshop evaluation sheets (**WES**) were reported with descriptive statistics to identify the frequency of the 20 statements related to assess participants' perceptions in relation to which extent they agree or disagree with the objectives of the workshop, contents and materials (Part A). These items were based on a Likert scale from 1 meaning "*strongly disagree*" to 5 meaning "*strongly agree*". The qualitative aspect took into consideration two features; the part B of the WES which included 5 open-ended responses which were scrutinised through a thematic analyses among educational sectors and the focus group interview (FG) to support the comments written from the WES. Table 34 presents a statistical summary of the WES at item level with the features

of minimum and maximum values based on the Likert-scale from 1 to 5, as well as the mean (*M*) and standard deviation (*SD*).

**Table 34 Workshop Evaluation Sheet (WES) Analysis**

Statements	Min	Max	M	SD
1. The objectives of the training were clearly cover	5	5	5	0
2. The objectives of the tasks were related to the contents of the course	5	5	5	0
3. The content of the course was relevant to my teaching practice	4	5	4.7	0.48
4. The content was structured and delivered adequately	4	5	4.8	0.42
5. The materials distributed were helpful	4	5	4.9	0.32
6. The materials were connected to the discussions and presentation of content	4	5	4.8	0.42
7. The materials covered too much they were not clear enough	1	3	2.3	0.82
8. The contextualized course met the current needs in my institution	3	5	4.6	0.84
9. Participation and discussion were encouraged	4	5	4.7	0.48
10. The tasks were explained in detail to obtain better outcomes.	4	5	4.7	0.48
11. The list of references was useful to understand the structure of the course.	3	4	4.2	0.79
12. Presentations, videos, tasks and journals were beneficial for my learning.	4	5	4.8	0.42
13. The instructor used a different materials and resources to explain the subject.	5	5	5	0
14. The instructor stimulated my interests to complete the course	4	5	4.9	0.32
15. The instructor developed a large number of activities	4	5	4.9	0.42
16. The instructor provided materials to read and be prepared for each session.	3	5	4.7	0.67
17. The instructor motivated me throughout a diversity of tasks and activities	5	5	5	0
18. The length of time for instruction and tasks was enough	3	5	4.2	0.79
19. The course was useful for my professional development	4	5	4.8	0.42
20. the tasks were suitable for the purpose of professional development	4	5	4.8	0.42

The first part was in relation to the objectives and the contents of the workshop (items 1-4). All participants considered that the objectives were met ( $M=5$ ) and the contents were appropriate for empowering participants' assessment knowledge, skills and practices. All participants pointed out they agreed that the contents were structured and well delivered by using different resources by instructor applied (e.g., learning portfolio, power-point presentations, videos, etc.). Participants also stated that the contents were related to what they were currently experiencing in their classrooms, with some participants from the private and language institute sectors opining the contrary with the section of item analysis, and that was an opportunity to explore different areas of language assessment which were not seen in their previous academic programmes nor teaching qualifications and training.

*'The contents were very essential for my teaching practice and how to improve the testing resources in the classroom' & 'The course contents were very useful because I can identify how to plan and design my test better' ([WES] PUB)*

*'The sessions were very complete and they were interesting' & 'I think I learnt a lot because the contents of the sessions were accurate for my practice' ([WES] PRIV)*

*'The contents were very complete with theory, practice and how to incorporate them into my teaching practice' & 'I feel confident because I was able to revise what I knew and learnt a lot ...' ([WES] LI)*

These results suggest probable growth in participants' confidence and knowledge in the field. Findings from the focus group (FG) interview also highlighted that participants from the

public sector agreed the course was *'useful'* and *'interesting'* for their teaching profession in general, to improve their assessment skills and practices. Participants discussed of the opportunity that assessment could bring to their classroom beyond students performances and tests *'...definitely this course was useful for making important decisions, not only about my students' scores but how I should amend my teaching to seek better opportunities to enhance learning'* ([FG] PUB-T2). This participant also indicated that this workshop should have been given at the beginning stage for pre-service programmes since it was relevant and could have brought better opportunities to be aware of the use of resources to improve the quality of teaching and learning practices through the use of tests and assessments. Participants from the private sector pointed out that the workshop was very stimulating to improve their assessment skills and practices and highlighted that accurate assessment knowledge was important regardless teachers developing or not their tests. Similarly, participants from the language institute said that the course was contextualised and oriented to Tamaulipas teachers' needs and it provided an opportunity for teachers across their educational institutions to discuss different methods of assessment and techniques to evaluate students' language proficiency.

*'I believe that the course was very complete and useful, I wish I could have learnt this in my degree or TKT'* ([FG] PUB-T2)

*'This was a hard-working course in which I learnt a lot about my teaching practice, not only assessment but it was a great aspect to realise how assessment can help instruction'* ([FG] PRIV-T2)

*"This is the first time I take a course that actually is related to how Mexican teachers actually work. It was a very interesting and relevant with our needs"* ([FG] LI-T2)

The materials used during the workshop were evaluated and discussed by participants in the WES and FG. Four participants (two language institute, one public and one private) indicated that the use of a folder with all the printed power point presentations developed for the workshop were efficient since participants were able to revise the contents they were learning at the same time. Two participants (both from the public sector) pointed out that they would like to have more readings and more materials in which participants could expand their knowledge. All participants agreed that the number of hours and the number of tasks were appropriate for their personal time. Participants also stressed that group discussions of several themes related to assessments were beneficial to share ideas with other teachers to tackle assessment issues in their institutions.

In relation to the practical and reflective tasks, participants claimed they have made progress and developed their interests in the practical tasks. They pointed out that the practical tasks were *"appropriate"* and *"new"* to their teaching and assessing practices ( $M=4.7$ ). Participants strongly agreed that the purpose of the tasks were related to the contents of the

course ( $M=5$ ); informing that the course contents were coherent and contextualised. Five participants (three language institutes and two public) indicated that the *development of the task specifications* and *item analysis* were completely new to their knowledge and they claimed they have expanded their assessment skills with regards to the quality of new testing instruments (measurement theory) they aimed to develop. Besides this, participants mentioned that the reflective tasks were, to a certain degree, contextualised and they were useful for adhering what participants' currently required in order to conduct adequate tests and assessments in the classroom.

*'The materials were appropriate and the presentations included different examples and theory' & 'The portfolio included many activities and possibilities to revise what we learnt' ([WES] PUB)*

*'The item analysis was difficult and it takes time but it useful for those who develop tests' & 'Possibly we could have more readings but this could be a bit more extensive' ([WES] PRIV)*

*'The materials were great for me, the instructor selected adequate materials for my learning style' & 'The tasks were accurate to my needs and they were very efficient to expand my knowledge' ([WES] LI)*

Participants mentioned that all practical and reflective tasks were relevant to their teaching context, but they showed more interest and awareness towards the importance of item analysis. Participants from the private sector pointed out that they would require more time to actually conduct this type of analysis, but it could be time consuming. Also, this cohort reported that evaluation of items did not have sufficient time for the in the reflective tasks. They were also interested in conducting item analysis, but PRIV-T1 said that this was not a task for teachers, but more for those authorities who developed and delivered the tests. Similar findings were also reported in the analysis of the tasks among participants from the private sector.

A participant from the public sector (PUB-T3) shared her experience in conducting this analysis during the workshop and she indicated that some items in her previous tests were straightforward. She claimed that the workshop assisted to understand the scores of the students since this participant believed that the lack of preparation to develop a test might affect the students' performances. All participants agreed that the reflective task 2 was reflection of the practical task 1, which had more impact on participants to explore that assessment did not only represent a test and a mark, but that it goes beyond the classroom. Identifying their assessment practices, skills and the possible issues of malpractice let participants become much conscious about technical aspects in language testing and assessment. The reflective task 3 was outstanding and, to some extent, relevant to the participants since this was in relation to evaluate something they had been struggling with a particular aspect of assessment. PRIV-T2 also highlighted that this task related to the problems and solutions was the most applicable to the current teaching

practice. Data reported that participants from language institutes were equally content with the reflective and practical tasks.

*'...when I was developing the test and evaluating the items I was able to understand that I possibly design my tests very easy to help the learners. I now can identify if the items were difficult or easy for my students. We...teachers...we need to know how to do this .... ([FG] PUB-T3)*

*'...no for us, I don't think I would do item analysis if I am not even develop my tests. I hope my coordinator does that with every test... teachers don't have time to teach, mark and evaluate items.' ([FG] PRIV-T1)*

*'...very relevant tasks, number 2 was difficult but it opened my eyes to know that the test and a number are only about assessment. Everything is assessment in the classroom' ([FG] PRIV-T2)*

The participants agreed that the tasks were valuable since they were able to reflect on their current assessment needs and the critical issues teachers tend to face daily, but participants also expressed that more time for practicing should be required in order to expand their assessment skills and practices. Participants claimed to have improved their assessment skills and practices, not because of the contents or tasks during the workshop, in fact because they had opportunities to discuss, evaluate, reflect and practice different assessment features throughout the tasks that were conducted in the workshop.

Participants suggested that this workshop should be available not only for teachers but for other stakeholders such as language programme managers and pre-service teachers). One participant indicated that this could be accessible for those teachers who are being trained to become professionals in the field of language teaching. Two participants (one private and one language institute) showed their interest to explore other topics concerning the use of assessment using technology.

*'It was a very nice opportunity to expand our knowledge and use different aspects to improve the assessment in the classroom' & 'This should be available for the ELT community in Tamaulipas and possible for other educational levels such as primary and secondary schools' ([WES] PUB)*

*'This could be great for language programme managers since they are in charge of developing language tests' ([WES] PRIV)*

*'It was a great opportunity to learn but I would have loved to explore the use of assessment in the computer' ([WES] LI)*

All participants mentioned that they would like to recommend the course to other teachers because of the contents and the free access to the materials. PUB-T1 indicated that the course should be delivered to teachers who usually work at basic and secondary education programmes due to the importance of assessing the English language at an early stage. The six participants also highlighted that this workshop could bring opportunities for inexperienced and experienced teachers who are involved with assessment.

*'Something I really enjoyed learning was to know other methods to assess students; especially methods in which students can reflect themselves such as self-assessment and develop rubrics' ([FG] LI-T2)*

*'...all teachers, no matter their discipline or course they are teaching, should take a module on evaluation, this is very important because we need to make decisions not about our students progress, our progress is also important' ([FG] LI-T3)*

In relation to the instructor's performance, comments were written by the participants. Findings showed that the instructor could have given more materials to discuss and practice more during the workshop, but they also showed that the time and materials adequate for the duration of the workshop. One participant from the language institute sector mentioned that the instructor should have provided extra hours for further guidance.

*'The instructor was very patient with my doubts and explained very well each of the course contents' and 'The instructor should have provided extra time to explained the results of the practical tasks' ([WES] PUB)*

Participants indicated that the length of the course was enough, but they would like to attend to further trainings in which they could practice specific areas of language assessment. For instance, participants from the public sector indicated they would like to know more about the test development processes due to the fact they were constantly developing worksheets and small tests for their students. PUB-T2 pointed out that the course could have more opportunities to share different tests and compare the usefulness, especially test validity, reliability and practicality. Participants from the private sector mentioned that they would like to work more with marking criteria, especially for writing. Moreover, LI2 participants pointed out that the use of rubrics, peer- and self-assessment were vital since in their institutions the aspect of autonomy plays an important role. A participant from the public sector indicated that the course should have included more practical aspects rather than theory. This informs that participants require a larger number of hours for their professional interest.

Participants were also asked about the last session of the workshop in relation to the evaluation in the Mexican State of Tamaulipas. Participants found relevant the historical overview of how the English language was introduced in the Mexican educational system at the end of the 1920s and how this had evolved during the last three decades. Participants indicated that English has been in constant changing based on the SEP and the government.

*'...I didn't know most about the evolution of the English language. I was supposed when I knew that English started in the 1920s for the private sector...' ([FG] PUB-T2)*

*'...A lot of educational reforms have had the attention of the government but these have not had the impact they required...' ([FG] PRIV-T1)*

Also, participants highlighted that assessment in Mexico was more oriented to demonstrate what you know without the aspect of development. Two participants had a discussion about the

difference of the words 'evaluation' and 'assessment'. They highlighted that assessment have no a translation in participants' first language (Spanish). This was an interesting fact among the participants due to the essential and vital features that assessment provides.

The findings from WES demonstrated that participants were comfortable with the workshop in relation to the contents and objectives, practical and reflective tasks, materials and resources, practicality, and with the way in which the instructor (the researcher) performed during the instruction phase. Participants showed positive attitudes throughout and highlighted that the workshop was useful and beneficial for their professional development as well as to empower their knowledge, technical skills, their classroom experiences and the relevance of assessment towards the Mexican State of Tamaulipas. The following section of this Chapter reports on the impact of the workshop towards participants.

## 6.4 Impact of the Workshop

Participants were asked to identify whether the workshop might have had an impact on their teaching, learning and assessing language practices. Four main themes arose: a) participants' practicum trajectories, b) developmental trajectories, c) institutional operational assessment systems and d) local and national considerations for improving assessment skills and practices.

### 6.4.1 Participants' Practicum Trajectories

In relation to possible impact on participants' practicum trajectories, five participants indicated that the workshop was helpful, to some extent for their students' cohorts. This was mainly because participants demonstrated they have developed confidence in using different assessment approaches and methods in their classes. Two participants (PUB-T2 and LI-T2) highlighted that they felt more self-confident in providing feedback and use error correction techniques in the classrooms where the number of students is larger. All participants pointed out that their way of delivering feedback has slightly changed since they now believe the focus was on what the learners need in relation to the learning objectives of the course participants were teaching at the time. Similarly, four participants also highlight that the workshop was useful to be able to improve the assessing methods for students' speaking and writing skills. These participants also indicated that they had developed testing resources which were much aligned with their current teaching courses and learning objectives.

*'...after the workshop, I believe I have been more conscious on how to deliver feedback and distribute my time in class with those ones who require more teaching...also, how to fix students' mistakes especially in speaking and writing' ([FG] LI-T2)*

Two participants (PRIV-T2 and LI-T3) discussed that they felt more confident in relation to students' guidance with their formative assessments, the use of learning portfolios, and rubrics for productive skills. Moreover, participants discussed that the workshop helped them understanding different types of items and tasks, so they could provide adequate guidance to their students to answer particular items (e.g., fill in the gaps, information items, etc...) for reading and listening tasks. Some participants indicated that some of their students were not in their adequate language learning stage and this was mainly because of their previous teachers' assessment practices or the misuse of affective assessments (e.g., attendance, homework and participation). Three participants (PUB-T1, PRIV-T1 and LI2) also highlighted that the workshop was appropriate for teachers to keep relevant students' academic records; not only students mark, but to track their learning progressions. In addition, these previous participants also indicated that the tasks from the workshop were suitable for their needs compared to the way other institutions tend to deal with assessment practices, and to which extent the malpractice and misuses could affect students' performances.

*'...something I found useful was to develop a way to keep my students' records, we sometime believe that students who arrive to a new level have the same level of English. Reality is other... it is difficult to teach a class with different levels. I was able to realise the importance of having students' records in class and the benefits of this towards learning' ([FG] PUB-T1)*

*'...been teaching English for almost five years and never had an intense course of assessment. Useful to identify what my students needs to develop a record of activities much more oriented to the class and avoid irregularities...' ([FG] PRIV-T1)*

The workshop has allowed teachers to be slightly more critical in selecting and using other materials for their assessment practices. One participant commented that standardised tests were used for different institutions to highlight the quality of teachers' instruction, but they did not improve nor demonstrate how much knowledge students have acquired in those courses they were currently teaching. One reason that this participant mentioned was that students should not rely solely on their classroom assessments, but teachers should look opportunities to boost students' potential by teaching them how to reflect on their own learning (use of autonomy). The rest of the participants agreed that self-assessment was important, but other assessment methods could also complement students' learning such as tests and alternative assessments. These findings demonstrated that participants have become much aware of the use of assessments in their language classrooms. The following section reports on the outcomes at the professional development trajectories of the participants.

### 6.4.2 Participants' Developmental Trajectories

Participants also mentioned that they have become more conscious on how they currently perceive assessment as well as to identify the qualities and irregularities in conducting a range of skills and practices in their classrooms. Participants from the public sector indicated that the methods they had conducted assessments before the workshop had to some extent not met the learning objectives and were less orientated to what students actually required to learn '*...the course has help me be aware that I my tests weren't that bad, but not really coherent, they were easy and not really realistic..*' ([FG] PUB-T1). This participant also showed interest in developing task specifications for her summative tests. The private sector also recognised the opportunities that assessments could provide a large number of advantages for the development of the teaching methods and practices, having as an example '*...I want to develop tests because I will be able to know if my teaching is worth it, the test from XXXXX are just a copy of the exercise with examples from London' ...*' ([FG] PRIV-T2). This participant claims that by developing language tests, she would be able to recognise how well or bad her teaching practice was. Participant LI-T3 indicated that she had developed a criteria to revise their students' writings, '*what I found interesting was the way to provide feedback, so based on the contents of the XXXX I developed a rubric (marking scheme) to guide my students and me...'*'. These were examples of participants' developmental trajectory in relation to awareness in assessment.

Participants also admitted that they expanded their knowledge in different areas of assessment including the contents from the syllabus of the workshop, and that three others emerged: a) self- and peer- assessments, b) historical evolution of the Mexican evaluation system and c) assessments as educational support. Participants showed their enjoyment especially to realise how Mexican teachers tended to rely solely on tests, this refers to the testing culture, rather the assessment culture. A participant from the public sector mentioned that even the formative tasks were summative in the end '*...everything students do in class is being assessed formative and summative, we have no difference...'*' ([FG] PUB-T2). Developing resources that students could self-evaluate their own progress was a factor that the private sector and language institute highlighted as essential and needed to be developed. Critical awareness among participants raised due to the importance of the use of alternative assessments to empowering learning after the workshop. Participants indicated that they learnt how to use different techniques for portfolios and marking productive skills, but at the same time how tests and assessments resources should rely on theoretical and practical foundations to maximise their efficiency. Participants also highlighted that the use of portfolio for students was really relevant; however, they also claimed that teachers should have a 'teaching' portfolio where they can identify tasks based on their experiences, and know how to improve them to strengthen their

learning practices in the future. Materials from this workshop were given in a folder to participants who usually referred it to as portfolio.

Something that got researchers' attention was a discussion during the focus group in which two participants from the public sector agreed that professional development courses and trainings they have attended did not meet their expectations *'focused on what teachers should assess, whereas it should be on how teachers should assess'*. Likewise, LI-T2 and LI-T3 pointed out that *'teachers should not be afraid to explore assessments, because it is evaluation, and we are evaluating everyday'* ([FG] LI-T3). *'...we, as teachers, need to be trained in all professional development aspects equally...'* ([FG] LI-T2). These two participants perceived assessment as an essential aspect teachers should know because the teaching methods, skills and practices are in constant change, teachers need to be critical and this arises when teachers assess. These findings showed that participants acquired more confidence, knowledge and awareness as a result of the workshop. The following section reports on the outcomes on how the workshop impacted on participants' institutional operational assessment systems.

#### **6.4.3 Participants' Institutional Operational Assessment Systems**

The workshop was useful and practical to bring updated and useful assessment skills and practices into the language classroom and institution. A participant claimed that the workshop was an opportunity to share ideas through reflections to promote better assessment practices among institutions. All participants pointed out that assessment was sometimes misled because of the way it normally takes place in their institutions. For instance, participants, mainly from the public sector and one from the private, indicated that they felt confident in conducting assessments, but they have a main restriction which was the lack of time to prepare and plan materials for assessing students' performances. In addition, PUB-T2 highlighted that the number of students in the classroom did not facilitate the process of delivering adequate feedback; on the other hand, PUB-T1 commented that she had delivered feedback in groups, she made a reference to one of the techniques she learnt in the workshop.

*'I think something that the workshop helped me a lot was to give feedback to my students. I have 50 students and it was so difficult for me to provide individual feedback....'* ([FG] PUB-T2)

*'...Yes, I understand you completely, what I did was to group my students based on similar English level and mistakes they had made, it was a better way and they have been quite surprised I was able to handle it...'* ([FG] PUB-T1)

Differently, participants from the language institute and the private sector were more focused on alternative assessments methods because they were not responsible for developing their language tests. PRIV-T2 indicated that he had recently been hired in an institution where tests needed to be developed, and mentioned that the course was suitable for this since he gained experience to be more confident and develop better testing resources, as well as to interpret different frameworks. By the same token, LI-T1 stated that she is more aware and conscious on how to check students' progress.

*'...the most useful part of this course was to analyse tests and items, I just got a position of English coordinator [language programme manager] and I am responsible for developing assessments, I am so happy I was able to take this course and for free' ([FG] PRIV-T2)*

*'...I thought that developing tests was easy since I used to copy items from internet and website, but I now feel more confident on the way I select items and deliver feedback to my students...' ([FG] PUB-T2)*

Moreover, participants highlight that the institutional authorities have been the main barriers to be able to develop participants' confidence, knowledge, skills and practices. The restrictions from their institutional operational assessment systems, mainly in the private sector and language institutions, might lead to lower levels of teachers' assessment literacy. LI-T2 indicated that teachers should explore other opportunities to be able to identify how to improve assessments, not only for their students, but also for their practices and institutions. Public and language institute sectors stressed that some teachers have not been well trained in assessment and they passed their students because of other attributions (e.g., students' participation in class, homework, attendance, etc...). The public sector pointed out that they assess students' effort, but these were not that significant towards students' final mark. Similarly, private and language institute sector pointed out that the abovementioned aspects were not included in their institutional criteria. Participants also expressed their content with the methods on how to deal with certain students' affective features. All participants mentioned that these aspects were important to include as part of their assessments. However, LI-T2 pointed out that teachers should be aware to which extent they were helping their students to pass, as an example, *'... by passing students just because they were motivated and participative, does not secure a successful learning journey path...'* ([FG] LI-T2). Tensions among teachers and language programme managers were also reported from the findings of the workshop tasks. The following section reports on the findings from a possible impact beyond the classrooms and institutions.

#### **6.4.4 Participants' Local and National Considerations**

Participants discussed two emerging features for the impact at socio-political level: a) English language teaching programmes at upper secondary schools and higher education

## Chapter 6

institutions, and b) the assessment needs for educational programmes to train EFL teachers. Participants pointed out that there was not a clear national curriculum for English language learning for upper secondary schools and higher education institutions. PUB-T1 and PUB-T2 highlighted that the government has made a lot of effort to implement English language programmes for basic education, but not for higher educational levels where English should be a priority.

Also, participants indicated that the federal government had not provided enough support for institutions, especially for higher education institutions. PRIV-T2 pointed out that the Ministry of Education (SEP) and several universities oblige students to take English qualifications as Exit Requirements, but students tend not to be well trained, especially for the public sector. PUB2 also mentioned that in his institution there were just two compulsory English courses, but these were not enough, according to this participants' experiences. A Participant also indicated that this sector has been in constant training to improve teaching quality, but there was a lack of national support for teachers. Participants also discussed the current national reforms of English, they mentioned that not many teachers have higher levels of language proficiency (C1 or C2 CEFRL) and these certifications were difficult to achieve.

*' the government and the universities should work more to implement better reforms and support teachers to expand their language courses so as to meet the objectives from the national curriculum. At XXXXX there are just two English courses with no more than 60 hours each, this is not enough to get a B2 for graduation purposes...'* ([FG] PUB-T1)

*'...when the SEP wants to implement something, they required higher levels of proficiency and examinations from Cambridge which are very expensive and not all teachers have opportunities to pay that...'* ([FG] PUB-T2)

All participants agreed that the workshop was relevant for their teaching practice and this could be an opportunity to collect evidences on how the Mexican educational systems were formed. Participants highlighted that the public sector requires much more work in relation to their infrastructure of schools and the number of public schools available. Also, participants suggested that the Federal government has implemented several programmes for higher education, but there has not been a significant change. Participants from the language institutes pointed out that they have been working as autonomous institutions from the Federal government, and that there should be a link with language institutes that have been working exhaustively to improve the quality of the public and private sectors.

### 6.5 Summary of the Chapter

This Chapter reported the findings to answer the research question 4 that dealt with participants' progress in the workshop and research question 5 that dealt with measuring the

impact of the workshop on the participants through a focus group interview. This concurrent and mixed-method research approach included the pre-measure and post-measure findings, the reflective and practical tasks, the workshop evaluation sheet and the focus group interview.

The pre-measure and post-measure were vital to determine if there was any progress among the participants, but also to identify which assessment themes participants were more or less interested in. Although this tool has a limitation and requires further revisions, it was used as a guideline to identify different participants' trajectories. Data reported progress for all candidates, but a higher discrepancy was captured within the public and private sectors. From the post-measure findings, participants showed they became skilled and developed their confidence and awareness in different assessment components such as item analysis, types of tests and items, development of tests, concepts of assessments, communication of assessments, alternative assessments and assessment as educational support. All these components were in relation to both technical considerations and school-based assessment considerations.

For the public sector, participants indicated that they made more progress in areas such as item analysis, types of tests and items, development of testing instruments and alternative assessments. For the private sector, there were similar results from the public, but they also indicated more progress noted in educational assessment as a support tool and testing principles. Lastly, the language institute had similar responses as the previous sectors being the areas with more improvement in item analysis, types of items and tests, testing principles and development of tests. These participants did not reflect any progress in marking schemes since this demonstrates that participants were aware of the uses of holistic and analytic rubrics for assessing the productive skills.

The practical and reflective tasks provided an in-depth analysis of the current status of language assessment literacy, but also these demonstrate where participants had more interest to improve their professionalism. These three reflections and two practical tasks were key elements to enable teachers to empower their professional development. It is important to remember that professional development opportunities are the ones where teachers can reflect, work with other teachers and to provide solutions on area to improve in their educational settings. Each participant was able to learn, revise, reflect on a total of 10 different sessions: 1) introduction to the workshop, 2) testing, assessment and evaluation in language education, 3) language testing, 4) language assessment standards, 5) language assessment in receptive skills, 6) evaluation of items and test, 7) language assessment in productive skills, 8) communication of assessments and error correction, 9) alternative assessments, and 10) language testing in Mexico and Tamaulipas.

The findings from the workshop evaluation sheet and the focus group in relation to the satisfaction of the workshop were not only to demonstrate the positive attitudes from the participants towards the use and instruction of the workshop and materials, in fact these can also be used to empower the use of other professional development in other areas where Tamaulipas EFL teachers require further improvement. These findings also indicated that participants were satisfied with the instruction of the contextualised workshop. Participants reported that the objectives of the course, as well as the tasks were met. The materials and resources were adequate, and participants showed interest in different assessment contents and practices. Participants indicated that an extension of the training would be beneficial for them. Also, some suggestions were made by the participants in relation to including more practical tasks than theoretical features of the field. Moreover, participants also highlighted that the contextualised session (Language assessment in Mexico) of the workshop was relevant because participants became much more aware of the current situation of evaluation in Mexico. Overall, participants showed positive attitudes towards the course and found it useful for their teaching practice.

The focus group interview allowed participants to discuss whether the workshop had had an impact for their personal careers and their institutions. There were four themes: 1) participants' practicum trajectories, 2) developmental trajectories, 3) institutional operational systems and 4) socio-political aspects in view of local and national considerations. For the participant's practicum trajectories, the workshop assisted to increase awareness of the different aspects of the sessions, but also they were able to put in practice what they learnt and revise during the sessions. The workshop also assisted participants in reflecting and working in collaboration to with other colleagues from the same educational sector and other educational sectors to be aware of the differences of language assessment. Professional development was achieved among these participants due to the opportunities in analysing and observing differences and similarities. However, several tensions among the participants with their institutions were noticed; the lack of support for professional development opportunities and the lack of being taken into consideration for test development, could indicate that institutions should be open more to work more in collaboration. Finally, institutions are now taking more into consideration the assessment practices from the teachers, this workshop and more professional development opportunities could be adapted to empower the national system of English language teachers.

## Chapter 7 Discussions and Conclusions

This research study followed two types of methodologies. On one hand, an exploratory research was conducted to analyse the status quo of EFL teachers' language assessment literacy, to discover discrepancies among three educational sectors and to investigate teachers' backgrounds in professional development and language testing and assessment. On the other hand, an experimental study was carried out through a workshop to measure participants' influence and progress in relation to teachers' language assessment skills and practices and their impact towards their language assessment literacy and professional development on EFL teachers in Tamaulipas. A total of five research questions were addressed based on the research aims, with mixed-method and concurrent mixed-methods approaches which were applied to analyse four phases of data collection: 1) Survey (TALDI), 2) Semi-structured interviews, 3) workshop activities: pre- and post-measures, practical and reflective tasks and a workshop evaluation sheet, and 4) a focus group interview. Three educational sectors that were explored: the public, the private and the language institute sectors.

This Chapter presents a discussion of the findings to answer the five research questions for this study. The first section provides a discussion of the first research question that deals with the status quo of language assessment literacy in EFL teachers in Tamaulipas. The second section gives a discussion on the main differences and similarities comparing the three educational sectors investigated regarding the teachers' language assessment skills and practices. The third section discusses the relationships between teachers' professional development and language assessment literacy backgrounds. The fourth section presents a discussion of the EFL teachers' progress and language assessment areas which participants considered essential for the purpose of assessment in the language classrooms in Tamaulipas. Then, the fifth section includes a discussion of the impact from the workshop on the participants. The following section discusses the contributions towards the knowledge of language assessment literacy, professional development and methodological decisions. Then, the limitations of the study are presented and discussed. Finally, this Chapter concludes with a section of further opportunities to continue exploring the field of language assessment, not only in Mexico but worldwide.

### 7.1 Status Quo of Teachers' Language Assessment Literacy in Tamaulipas

This sub-section discusses the findings to answer research question 1 (*what is the status quo of language assessment literacy of EFL teachers in Tamaulipas?*). This discusses the definition of assessment literacy to present the findings that status quo refers to: 1) test relevance and

occurrences, 2) test responsibilities, and 3) assessment of language skills and linguistic components.

### **7.1.1 Test Relevance and Occurrences**

This research study followed Fulcher's (2012) definition of assessment literacy which has been, up-to-now, the most pertinent due to the fact that it has four main components: practices that include knowledge, skills; principles and concepts dealing with assessment processes to guide teachers to improve classroom practises, and finally, considered the most relevant for the purpose of this study, the assessment context that deals with historical, social and political reasons and impacts (2012:126). This investigation explored three different contexts: the individual (teachers), institutional (language programme managers) and the educational sectors of EFL teachers in Tamaulipas, Mexico

Mexican education relies primarily on academic achievement (Martinez Rizo, 2012) this means that the culture is towards testing; rather than assessment (Inbar-Lourie, 2008a). In other words, the importance of evaluation across the Mexican educational levels is relatively high due to the outcome of standardise tests (national measures) or classroom tests. Findings from Phase 1 (Survey: TALDI) and Phase 2 (Semi-structured interviews) demonstrated that EFL teachers in Tamaulipas used these tools to make critical educational decisions such as to upgrade students to higher proficiency levels, to identify students' English current language ability and to promote learning by marking and providing feedback to students. All participants indicated that the use of tests was essential for their institutional intentions. In addition, findings from the semi-structured interviews suggested that participants mainly relied on these tests for making judgements and educational decisions such as giving learners opportunities to discover areas for academic improvement, diagnosing the status of learners' language learning trajectories and for institutional matters such as teachers' performances and evaluation of educational needs. EFL teachers, as well as language programme managers In Tamaulipas found assessments relevant to different educational aspects: the use of assessments towards learning, empowering teaching practices, practitioners' professional development and institutional assessment practices. These findings were similar to the study conducted by Muñoz, Palacio, and Escobar (2011) within the context of Colombian EFL teachers. Also, participants pointed out that the use of appropriate, well-developed tests and classroom assessments could bring better teaching practices for their teaching space so as to possibly increase students' language learning efficiency.

Tests and assessments have demonstrated that they could be powerful tools to assist in the process of evaluation, selection, mobility and education (Alderson, Clapham & Wall, 1995; Martinez-Rizo & Blanco, 2010). The findings for this research question suggest that participants

had positive perspectives towards the field of language testing and classroom assessment. The majority of the participants indicated that they were confident, and to some extent, prepared to conduct tests and other assessment practices in the classroom; similar findings concurrent to the ones conducted by Tsagari and Vogt (2017). These researchers also mentioned that assessments were important for their teaching practices due to the relevance they bring to students and language teachers (Muñoz, Palacio & Escobar, 2012). This suggests that EFL teachers in Tamaulipas administer language tests and standardised tests in their institutions so as to assess their students for both formative and summative purposes. These findings were similar to the research study conducted by Quilter and Gallini (2010) in which their participants became aware of the use of tests based on positive attitudes when assessment occurs towards other educational decisions.

Also, the qualitative data reported the number of tests administered in the classrooms in the state of Tamaulipas varies based on the educational level and sector. This research found that upper-secondary school teachers tended to administer more tests than higher education institutions. This discrepancy of the number of tests in relation to language assessment can be due to the norms and policies of the institution and the time of instruction from the teachers in the language classroom.

### **7.1.2 Test Development and Responsibilities**

In relation to test development, Brown (2012) and Fulcher (2012) indicated that for the purpose of improving learning inside the classroom, teachers should have in mind classroom assessments that are aligned with the course contents and how students are capable to demonstrate their linguistic competence. The findings reported that the majority of the participants from the public sector concurred that they developed their tests, whereas the private sector and language institutes specified that their tests were mainly developed by their institutional authorities or language programme managers. This clearly indicates that not all teachers have the opportunities to develop a language test. These findings imply that those participants who develop language tests may have advantages of using a diverse number of assessment skills and practices, which could lead to adequate language assessment literacy levels. Conversely, for those participants who omit the development language test process could suggest that they might lack the requisite skills in developing accurate and valid testing resources (Brookhart, 2001; Mertler, 2002).

The private sector and language institutes admitted that they use of standardised tests from international examinations in their classrooms to achieve institutional recognition and to identify how much students have been learning. These findings shared similarities from the semi-

structured interviews in which participants from these two cohorts used standardised tests to improve the quality of their teaching and prepare their students for these language qualifications. I strongly believe that the use of these tests could bring potential benefits to evaluate the overall performance of the students, but a lack of learning progression may not be reported nor observed. Teachers teach several contents, and the standardised test might not cover those features students were learning in the classes.

Also, the use of alternative assessment such as learning portfolios, worksheets, journals and essays were the most common tools for assessment resources, other than language tests. From the qualitative data, other external factors emerged, these were related to the assessment purposes, but not linked to language proficiency. In fact, they assess how students learn and perform during their language classes. The use of class participation, homework and attendance to the class were other features of educational assessment to empower the values and responsibilities from the students. These findings were also reported in studies from Rea-Dickens (2007), Ruiz-Esparza, Verdugo and Castillo (2012) in which indicated that the assessment practices were not only focus on students' performances, but also on how they performed.

Besides this, there was little evidence from the sectors in identifying ethical considerations or code of practice for language assessments. Only the language institute sector provided several guidelines for the administration of language tests. The other two sectors indicated that they used the national educational regulations from the Ministry of Education to administer tests and other assessments.

### **7.1.3 Assessment of Language Skills and Linguistic Features**

With regards to what and how EFL teachers in Tamaulipas assess their students, the three cohorts showed different perspectives for conducting assessments that includes assessing language skills (e.g., reading, listening, etc...), as well as, to which extent teachers considered essential for assessing students' performances. The findings of this research indicated that some institutions control what language features to assess and the methods or forms to assess language learners. As an example, the public sector stated that they assess language aspects such as grammar and vocabulary and sometimes they include items related to the reading and listening skills. , the private sector and language institutes showed a wide range of language skills and linguistic elements (e.g., pronunciation, grammar and vocabulary) Based on the qualitative data, this could lead to different reasons for assessing language such as the lack of time and the facility teacher have to assess these methods. Findings from the data triangulation denotes that assessments can also be treated based on other factors that influence the administration of tests in the classrooms the number of students in a classroom.

Moore (2017) and Ramirez-Romero, Pamplón and Cota Grijalva (2012) reported that Mexican teachers tend to assess grammatical structures or one or both receptive skills, but the productive skills were rarely assessed. These authors also claimed that this is due to the institutional influence from external features such as the number of hours for teaching, the teaching course level and for the discrepancy from students' language proficiency levels. The public sector highlighted that they did not assess speaking due to a lack of time for conducting this activity in the classroom. The language institutes showed that they assessed all language skills and they also included a section in their tests named as "Use of English" in which several items have been included to particularly assess students' grammar and vocabulary. Differences in language testing skills suggest that participants may have different levels of language assessment literacy and these levels could be influenced based on participants' educational sectors. These findings suggest that Tamaulipas is an example of the current situation Mexico is currently facing.

The qualitative data also showed three different operational assessment systems among the three educational sectors. An example of full operational assessment system, was mainly found in the public sector in which teachers were mainly in charge of the assessment practices, including test development. Restricted and partial operational assessments were found in the rest of the groups. The private sector tended to have a restricted system in which tests were given by the institutional authorities and teachers' classroom assessments were not considered for the purpose of summative assessment. Language institute showed a partial operational assessment system in which tests are delivered by the institutional authorities, but teachers are responsible to conduct other assessments to complement the students' final mark (Moore, 2017). These findings suggest that the assessment practices vary based on the educational systems in which teachers work for.

To answer this research question (What is the *status quo* of EFL teachers' language assessment literacy in Tamaulipas?), these findings suggest that Tamaulipas Teachers of English are attentive to the importance of the use of tests and assessment for their language classroom. Participants were aware of the advantages and disadvantages of the use of tests and types of items. A discrepancy between the number of tests per educational level and sector should be an aspect to explore and to analyse to measure whether this might benefit or affect students' language learning trajectories. The number of language tests should reflect on the structure of the length of course rather than multiple assessments without a focus (Jones, et al., 2005). Also, EFL teachers should be more involved in the test development since this could allow teachers to gain more knowledge and experiences at the moment of selecting which features they need for their language tests. In case the test development process relies solely on the school authorities, the language programme managers should take into consideration their teachers to allow them revise

and provide feedback, if relevant, to seek overall improvements of their institutional tests. The following section presents a discussion of the teachers' language assessment skills and practices and differences observed among the educational sectors.

## **7.2 Self-perceive Teachers' Language Assessment Skills and Practices (TLASP)**

This section aims at discussing the findings from the second research question (*What are the discrepancies, if any, in EFL teachers' language assessment skills and practices among different educational sectors in Tamaulipas?*). There were three components of Teachers' language assessment skills and practices (TLASP) which were examined in this research project: knowledge of assessment, technical considerations in assessment and school-based language assessments (see Figure 2). Findings from the Survey (TALDI) reported that participants were familiarised with their assessment skills and practices among different language assessment literacy standards. Moreover, statistical tests, Analysis of Variance (ANOVA), reported that there were significant differences in participants' assessment skills and practices among the three educational sectors (the public, the private and language institutes). In addition, the qualitative data (interviews) was also used to support these findings. The following section has been divided based on the three components.

### **7.2.1 Knowledge of Assessment**

Brookhart's (2011) review of educational standards to empower teachers' assessment literacy, knowledge and skills for assessing students' performances suggests that teachers should develop their skills whilst taking into consideration features of language awareness in relation to theoretical concerns. Maclellan (2004) and Xu and Brown (2016) also indicated that the knowledge of assessment was essential to develop skills and practices for effective usages of assessments in the classroom. There were three standards studied in this research (*Applied Linguistics, Educational Theories and Institutional Assessment Polices*) based on the conceptual framework of the literature. The TALDI survey reported descriptive statistics in that these three standards obtained slightly similar means for participants' assessments skills and practices.

Although, these numerical means were, to some extent, similar in TLASP, participants from the private sector had statistical difference in the standard of *Applied Linguistics*. This indicates that a difference regarding to participants' perceptions of language theories towards teaching in the classroom is presented. Findings suggests that their assessment skills and practices from the private sector were different from the public and the language institutes. TALDI data reported that some participants' academic trajectories in the private sector were not related to language

education disciplines which proposes that there were some multi-disciplinary teachers, which could indicate a reason for this statistical difference. Abell and Seigel (2011) and Shulman (1987) also considered that these “*multi-disciplinary*” teachers should strongly reinforce their skills and practices for language theories in relation to the contents of what they are teaching (e.g., applied linguistics to language teaching) to maximise their teaching practice and improve students’ performances.

For the standards of *Educational Theories* and *Institutional Assessment Policies* standards, findings from the survey demonstrated significant differences regarding participants’ skills and practices in language institutes than the public and private sectors. Participants from language institutes self-perceived that using and bringing a diversity of educational theories and practices into the language classroom improve the process of teaching-learning. This can be explained and supported from the interviews in that some teachers used different approaches and language skills not only to assess their students’ performances, but also to highlight efficient aspects for teaching improvements. From the qualitative perspective, interview findings indicated that the methods for assessing students from language institutes could lead to have better opportunities to manage a great deal of strategies that increase the institutional profile. This cohort largely informed that the assessment practices the four language skills, and an extra linguistic feature to assess grammar and vocabulary through the use of diagnostic, formative and summative purposes. TALDI also captured that the public and private sectors had no significant difference in their institutional assessment policies for these latter two standards. Moreover, the Survey and Interviews concurred that these two sectors followed the National Standards of Education proposed by the Ministry of Education. This could be interpreted based on the similar educational approaches for language learning and assessment from participants’ operational assessment systems.

Although this current research study did not consider any subject other than English, the educational sectors investigated recorded differences, similarly to the research study conducted by Zhang and Burry-Stock (2003). In other words, both these research studies concluded that the educational sectors’ decisions on learning, teaching and assessing language can be influenced based on teachers’ knowledge of assessment with regards to the three standards, professional trajectories, experiences and operational assessment systems. The following section presents the discussions of the findings from the standards of the component related to Technical Considerations in Assessment.

### 7.2.2 Technical Considerations in Assessment

For the standards of *Testing Principles*, the survey reported similar values in means that suggested participants were aware of the different test qualities and the elements to take into consideration the usefulness of tests, as well as to develop other testing resources. Data reported significant differences across the three educational sectors. The interview findings, on the other hand, reported that teachers tended to evaluate their testing resources by sharing with other colleagues. This is an example of face validity that was captured in the interviews with participants from the public sector and less often in the language institutes. Regarding to the private and language institutes, the identification of test properties relied on their language programme managers. These results advocate that participants from each sector perceived their assessment practices differently, regardless they understood the theoretical foundations, knowledge and principles of assessment. Perhaps not all teachers have access to analyse a language test based on the testing principles, but the application of the principles are embedded within the practice (Rodriguez & Zabala, 2016)

Findings from the standard of *Assessment Processes* suggest that EFL teachers in Tamaulipas have different skills and practices based on their institutional contexts. Some teachers demonstrated a lack of skills due to the fact they did not develop tests or they were not required to develop materials for assessing their students. Statistical differences were found across the three educational sectors. For the public sector, participants tended to have flexibility in deciding the appropriate methods for assessing their students, since several teachers can negotiate the number of tests applications. The private sector and language institutes showed that tests were given to their teachers to be administered and these participants might not possess adequate understanding in the skills and practices to develop testing resources. The lack of planning, design and pilot language tests could affect participants' assessment literacy. These findings complement the research conducted by Vogyt and Tsagary (2014) that some teachers do not have the same testing responsibilities in their teaching contexts, but awareness of the usefulness of the test could empower their testing materials (Bachman and Palmer, 1996)

The findings regarding *Assessment Purposes* suggest that participants were conscious of the use of certain testing materials, marking and identifying irregularities whilst conducting assessments (Sultana, 2019). These findings suggest that teachers are required to construct and administer language tests, but not all participants demonstrated adequate skills and practices for selecting the assessment purposes (Vogt and Tsagari, 2014). Participants' assessment skills in this standard indicated that the private sector had differences in identifying different properties of constructing items in relation to the construct. The qualitative and quantitative data in this study showed that the private sector teachers had limited access to develop materials due to their

operational system. In relation to practices, statistical differences were found among the three sectors. This can conclude that participants' practices varied based on how the way the curriculum is structured (Popham, 2006). Not all teaching contexts were similar, however, the process of knowing how to develop assessments for specific purposes could allow teachers enhancing their skills to identify and apply different assessment purposes in their classroom.

*Educational Measurement* standards obtained the lowest means for both skills and practices. However, statistical differences were reported mainly for the language institutes. These participants perceived themselves to have sufficient knowledge and skills from the other two sectors on how to identify and used different sources to analyse the tests and the quality assurance. In relation to the practices, differences were notable among the three sectors. The qualitative data reported that some teachers applied different approaches to measure the quality of tests and other assessment materials. This can be explained based on a lack of educational measurement training that was reported from the survey and the interviews (Tzagari, 2014). As an example in the interviews, the public sector reported that they conducted face validity, however the language programme managers from the language institute was the only participant who was knowledgeable regarding this. It is important to mention that none of the participants had actually conducted statistical analysis to identify whether the tests were useful, but they were aware of these possibilities (Malone, 2008; Fulcher, 2012). This also suggests that educational measurement cannot be that applicable in the Tamaulipas classroom contexts. The following section discusses the findings from the component of school-based language assessments.

### **7.2.3 School-Based Language Assessments**

The last component examined was the use of school-based language assessments in participants' classrooms. Four standards were examined: *Language Assessment*, *Assessment for Learning*, *Communication of Assessments* and *Assessment as Educational Support*. The findings showed that teachers have certain responsibilities in the process of selecting, conducting, marking, giving feedback and analysing test scores to improve the quality of instruction, and therefore learning. Davison (2004) and Scarino (2013) indicated that teachers' experiences play an important role in understanding and applying testing and assessment skills and practices in the classroom. The findings from this component showed that participants relied on their own perspectives based on how they were assessed while they were language learners.

Findings from the survey and interviews reflected that participants were skilled in conducting classroom assessments, which was reported under the standard of *Language Assessment* in relation to what teachers should teach based on the contents of the course (Yastibaş and Takkac, 2018). The public and private sectors showed statistical difference in terms

of assessing students' performances. The rationale behind this difference can be explained in the fact that these two educational sectors do not have the same opportunities to assess the four language skills as language institutes do. Quantitative and qualitative data obtained for this research showed that the public sector tended to have less time to assess their students and preferred testing techniques for reading and listening. This group had also mentioned that the flexibility of developing these tests can help them meet the requirements for marking them, but not recording students' progress and achievement in their language learning. On the other hand, the private sector uses different methods to assess language skills, but this cohort tended to rely on what the programme manager required to successfully meet the institutional standards, but they not focus on how students currently are learning. These findings were similar to the ones reported by the studies of Muñoz, Palacio and Escobar (2012, 2013), Giraldo (2018) and Giraldo and Murcia (2018).

The standard of *Assessment for Learning* showed significant difference in the public and private sector. It is important to highlight that this standard relied on alternative assessments and the frequency in which teachers provide constructive feedback to students to improve their language proficiency. This also concurred with the qualitative data which reported that teachers' previous experiences also influenced the selection of testing materials for the purpose of assessing students. Although, a statistical difference was reported in participants' skills for this standard under the two previous groups, the qualitative data reported that these two groups tended to follow national standards of education and have similarities in the use of alternatives of assessment. However, the operational assessment system from the participants of the public sector who demonstrated to have full access of assessments, contrasting to the private sector whose participants tended to have limited access, based on the institutional authorities.

The data collected reported that participants from the three sectors showed statistical differences on the administration of other assessment methods than language tests. The public sector followed more practical and easy-to-mark assessments methods like quizzes, worksheets and rubrics. The private sector, on the other hand, demonstrated an ample command of the language in terms of developing writing skills and linguistic features throughout essay, journals and group presentations. In addition, the programme manager from the private sector also commented that teachers should provide students with a great deal of learning resources to pass standardise tests such as TOEFL for academic and graduation purposes. For the language institutes, the use of portfolios played a significant role since most of the participants indicated that the use of learning portfolios may be misunderstood if teachers were not well-trained to use this tool in the classroom for the purpose of students' learning. Participants from this cohort highlighted that they use these tools to collect learning evidence, but these were not appropriate

revised or track for further amendments. This suggests that alternative assessments are also perceived differently from the participants, under the umbrella of assessment for learning. Arguably, other assessment factors emerged from the data. The public and the private sector also stated their concerns towards external assessments factors found out were related to how teachers assess learners' academic performances such as: responsibility of homework, attendance and behaviour. These factors were mainly found in the private sector.

I believe that Tamaulipas EFL programmes should be more conscious of the importance in conducting appropriate assessments regardless of their operational assessment access. Language teachers and programme managers are responsible in bringing quality to their learning and teaching practices, and assessment should be seen as an opportunity to identify and examine inaccuracies so as to maximise the learning process. Tamaulipas EFL teachers should be engaged to allow students to self-assessed and to demonstrate they also track how students projected themselves to reach learning goals (Brown & Abeywickrama, 2010) Teachers should change their perspectives towards the ample resources for using alternative assessments such as portfolios, worksheets, rubrics and journals to help improve the quality of assessment process.

For the standard of *Communication of Assessments* dealt with how teachers inform their learners and other stakeholders about the assessment methods and results. This research explored three different stages of communication: informative which teachers provide instructions to their learners of what they need to successfully achieve their performances. Instructional communication is when teachers provide feedback to learners and explore how to meet the learning objectives in the course. Finally, the communication of assessment is the last stage which teachers report on students' performances to their learners and other school authorities (Sethusha, 2014). For this standard, there was no significance on participants' assessment skills. They reported to be skilful in communicating assessment methods, results and inform other colleagues on how students have improved. Nevertheless, the assessment practices reported a significant difference from the public sector to private sector and language institutes. An explanation for this could lead to the fact that the public sector tends to have less regulations from the language programme managers and this could result to a lack of communication. Interview data also supported this statement since this sector has less opportunities to communicate or discuss assessments with other colleagues for different reasons such as the lack of confidence among the colleagues and that is less strict than the other two sectors.

Lastly, the last standard of this component was *Assessment as Educational Support*. This standard focused on how teachers should use language tests and other assessments scores to identify and improve the quality of instruction. The public sector reported a difference on how

tests are treated once students' evaluation has completed. This could be explained that the public sector has less opportunities to analyse and examine the test scores due to factors like lack of time or lack of knowledge on how to conduct this procedures (Grounds and Moore, 2017). A reason for this was found in the qualitative data in which these participants struggled with institutional issues such as the number of students and hours of teaching that do not allow to examine and explore further opportunities to use assessments to improve their teaching practice (Roux, 2018). On the other hand, in terms of practices, a significant difference was reported for the private sector. Once again, the operational system of these participants' educational sector is regulated by the school authorities which does not allow teachers to expand their skills to explore further improvement in their language classrooms.

Overall, there were differences among the three sectors in several standards. Assessment skills obtained higher averages than the practices, showing that participants are conscious of the skills they require and need to administer assessments, but a lack of assessment practices, from their institutions, is existing. Participants showed they were, to some extent, prepared in different aspects of assessment practices and skills that required for their educational contexts (Zhang and Burry-Stock, 2003) The communication of assessments was the only standard that did not showed differences; this allows us to state that EFL teachers in Tamaulipas respond to the currents needs to inform how assessments are structured and how students can achieve their goals. On the other hand, the three sectors obtained more differences in relation to the second component, technical considerations of assessments. This can be supported through the institutional operational assessment system. Tamaulipas EFL teachers' assessment flexibility could have led to gain experiences, skills and practices, but does not necessarily mean that they are assessment literate. The following section presents a discussion between professional development and language assessment.

### **7.3 Professional Development and Language Assessment Backgrounds**

This sub-section discusses the findings from the third research question (*In what ways has the workshop influenced EFL teachers' awareness of language assessment literacy?*) To begin with, Mertler (2009) indicated that some language teachers tended to underestimate their assessment skills and practices. This also concords with this study since there was a low percentage of participants (18.73%) who pointed out that they were not interested in the field of language testing and assessment, but they considered themselves to some extent prepared and self-assured with their skills and practices. Similar results were found in the qualitative data in which participants highlighted the importance of assessment as an educational practice, but other

factors such as the development of lesson plans and materials for teaching purpose were much relevant for their teaching practice than expanding their assessment literacy levels.

Data indicated that some of the participants had attended professional development opportunities as workshops or conferences, but it also showed a lack of training for participants in language assessment. Findings confirmed that language testing and assessment training opportunities were not considered essential for participants' teaching practice. The public sector demonstrated that assessment was an area of priority, but less priority was given from the private and language institutes. An absence of training for assessment from the participants was noticed since participants reported that they had not attended assessment trainings nor courses, but they considered that assessment was essential for their professional practice. Similar studies on the lack of training not only in language assessment, but in other areas in Tamaulipas was reported by Roux and Mendoza-Valladares (2014). These findings suggested that there is a lack of training of language assessment in the State of Tamaulipas to empower teachers for improving the efficiency of tests and classroom assessments. These findings matched with the majority of previous studies conducted by Brookhart (2011), Scarino (2013) and Inbar-Lourie (2013).

Berry, Sheehan and Munro (2017) reported that teachers' experiences could lead to make decisions on how to develop, administer and interpret tests results and other assessments. It is important to highlight that this study also indicates that teachers' experience as language learners also contributed to have an inclination on how teachers currently assess their language learners. These teachers' current experiences are constantly being modelled based on previous learning experiences, academic, and professional trajectories. This study indicated that EFL teachers and programme managers relied on their academic journeys to reflect and distinguish which assessment skills and practices to administer. Undergraduate and postgraduate programmes such as ELT/TESOL, Applied Linguistics and Elementary Education were the most notorious from the participants. A relevant finding was from the semi-structured interviews, in which participants showed their discontent to possess basic knowledge, techniques and practices to incorporate them into the language classroom from several teaching qualifications such as TKT (Teaching Knowledge Test) or ICELT (in-service Certificate of English Language Teaching).

Findings suggested that participants consider their academic and professional trajectories as engines to understand and apply a large deal of practices in their classroom. However, this study also suggested that these educational programmes, Bachelor's, Masters' and PhD's programmes, be limited to theory and practice. These findings were similar to McGee and Colby's (2014) and Mertler's (2009) studies in which participants lacked of training courses because they believed that their assessment practices in their classrooms were sufficient with the support of

their academic pathways, which they could have expanded their potential, but there was always room for improvements through professional development opportunities. Professional development could allow teachers to be trained with the basic foundations and with a contextualised focus on what teachers currently require to develop to promote in their language classroom better assessment skills and practices.

Data also reported a lack of motivation and support from teachers' educational institutions to undertake continuing professional development, not only in language assessment, but also in other areas. It is my belief that teachers claimed that what they had learnt during their academic journeys would be enough for their career. This lack of motivation to empower teachers' profile from their professional development comes directly from the institution where they work. Participants pointed out that a lack of support to improve teachers' professionalism is presented not only from their institutions, but also at national and federal level. Similar, this affected the aspect of language testing and assessment towards their implementation at local, regional and national standards (Martinez-Rizo, 2012). The qualitative and quantitative data collected from this study indicated that the feature of language testing and assessment is not a priority for Tamaulipas institution, but results showed that teachers are in fact, interested in developing this vital practice.

This research study followed the foundation that assessment is an educational practice that involves the integration of different agents such as language teachers and students in the classrooms, and language programme managers to assist teachers in their institution. The integration of these agents in the educational environment has showed that could facilitate better opportunities to improve the teaching and learning practices in the classrooms. Working collaboratively among teachers and language programme managers to improve the quality of assessments to maximise instruction could contribute to flag areas that requires current development. Mexico has a testing culture and tends to be remarkable since accountability tests have gained a lot of power to distinguish individuals' recognition. This recognition can be from students' achievement to the performance of the whole institution. In the context of education in Tamaulipas, several institutions have been relying on international and national standards to improve the quality of institutions. In addition, participants have stressed that there is a particular fear when assessments take place not only for students, but also when they are being observed and assessed by their students. The lack of experience and fear of being assessed might contribute to a lack of preparation and therefore this could affect students' performances. Besides this the lack of time from teachers since they can have full-time positions, motivation and incomprehensible materials have made difficult the process for teachers to be emerged in the field of assessment.

To answer this research question, there were clear implications to the field of language teachers' professional development to enhance their assessment literacy. The workshop was an example of a professional development to improve teachers' assessment skills and practices regardless their educational sector. The interaction between teachers from several educational sectors allow them to think, reflect and analyse how they are administrating assessments, providing feedback and examine current issues in language assessment in Tamaulipas EFL classrooms. I strongly believe that language teachers should be in constant training especially in the area of language testing and assessment due to the importance of this to improve the quality of educational system and sectors.

#### **7.4 Workshop: Areas of Improvement**

This sub-section discusses the findings to answer research questions 4 (*In what ways has the workshop influenced EFL teachers' awareness of language assessment literacy?*). This research question was answered based on the experimental research, taking into consideration the materials from the workshop and these were concurrent analysis with the focus group interview.

The areas in which participants indicated that they had learnt from the most were the assessment of receptive and productive skills, the use and application of alternative assessments for language classrooms, and critical issues in language testing and assessment in Tamaulipas classrooms. These findings also concurred with the study develop by Vogt and Tzagari (2014) Moreover, practical and reflective tasks showed that evaluation of language tests was also considered to increase teachers' awareness. They indicated that item analysis could be a time-consuming activity for them, but this might be a responsibility of the language programme managers or directors. However, participants from the public sector pointed out that this could be essential if teachers were responsible for one specific English level course, since these teacher tended to have multiple classes from different levels. Educational measurement could assist teachers in developing high-quality materials that they may reuse for further testing situations. Participants, besides this, showed progress based on the measures that was administered at the beginning and at the end of the course. Participants indicated they had learnt more about the different types of testing items to assess a particular language skill, as well, as different sources for assessing students without using tests (Leong, 2014).

Besides this, these findings showed that participants progressed in different aspects of language assessment literacy. Assessment skills and practices related to the development of tasks specification to improve the assessment purposes, the use of alternative assessments, assessment for learning to empower the students' summative assessments, and the evaluation of tests and items were the most aspects of improvement based on the data collected from the workshop

evaluation sheet, practical and reflective tasks, and the pre- and most measure. In addition, during the focus groups participants pointed out that this workshop had an impact on how teachers currently use assessment practices. There were other areas that participants did not make any progress such as marking schemes (analytic and holistic rubrics), this can be due to the fact, participants were already aware of the differences and how to apply these tools to assess productive skills. Nevertheless, they pointed out that even if they would like to incorporate new methods for assessment or strategies to empower assessment for learning, the institutional authorities would not appreciate their effort.

To sum up, this workshop in language testing and assessment was developed to explore and expand EFL teachers' language assessment literacy in Tamaulipas. This aimed at current needs that Tamaulipas EFL teachers highlighted important to have a positive impact on the developing of skills and practices to improve the quality of the assessments in different sectors. Besides this, it also demonstrated that participants were equipped with essential and practical tools to be applied in the language classroom to see benefits in students' learning trajectories. In addition, this course could be considered as an example of professional development opportunity for participants to reflect, gained knowledge and increase awareness of the opportunities that assessment brings to education (Berry, et al., 2017). The usefulness of conducting a workshop in language testing and assessment was in fact a feature to improve Tamaulipas teachers of English, but at the same time to explore in depth different tensions when assessment takes place among educational sectors.

### **7.5 Impact of the Workshop**

The last research question (*What was the impact of a workshop of language assessment literacy development on EFL Teachers?*) evaluated the consequential validity of the contextualised workshop to improve teachers' assessment literacy, but also which areas this workshop had a relevant impact. The data collected for the purpose of answering this research question was the focus group interview from the Phase 4. Data demonstrated that the contextualised workshop had four areas with plausible impact. These emerging themes were in relation to participants' students, their professional development, institutional impact and socio-political impact towards educational at the State level.

Although this research study did not include language learners to examine any direct impact, participants indicated certain awareness and confidence for adequate testing and assessment skills and practices that could contribute to an improvement of students' progresses in the classroom. To illustrate this, participants from the three sectors discussed that the use of alternative assessments such as the appropriate use of learning portfolios in class, essays and

journals, and how to take advantage of the use of students' workbooks, as well as worksheets for formative purposes. By the same token, the public sector also indicated that they believe that the summative aspects that not related to students' language performance such as class participation, attendance and homework could reflect a lack of students' achievements in relation to language performance. In addition, participants from the private sector discussed that they consider that the methods they have learnt and practiced from the contextualised workshop have helped students' learning journeys by delivering adequate feedback in classrooms where the number of students is larger. These examples showed plausible impact of the course towards participants' cohorts of students.

At teachers' professional development level, participants pointed out that they considered assertive with their assessment skills and practices after the course. They discussed that they gained "*knowledge*", "*techniques*" and "*skills*" that have helped them appreciate the purpose and management of the use of tests and assessment for educational purposes. This gives clear implications that the purpose of the workshop was met, and participants indicated they have developed certain abilities to improve and adapt new practices for their teaching practices. Some examples of areas of improvement discussed at the focus group interview were the understanding of different criteria for productive skills, development of language tests (assessment processes and purposes) and improve students' learning with assessments for learning. A factor that caused tension was that participants from TALDI showed that there was no difference in communicating test results; however, participants from the contextualised course showed positive attitudes towards the process of communicating and informing students about tests and their results.

At the institutional perspective, the focus group suggested that even if teachers were able to amend their assessment practices in their institutions, there would always be barriers from the school authorities. Data showed that participants from the public sector tend to have more control to make judgments in relation to their students; on the other hand, the private and language institute sectors have certainly less control since tests and assessments are regulated, to some extent, by their language programme managers. Participants indicated that they started developing materials and becoming much aware of different aspects when assessment can be misled such as: the number of students in the classroom, the distribution of students' English proficiency in a classroom and the hours of teaching and assessing language.

Participants highlighted that the workshop could also have impacted on the institution where they worked. Based on the participants' operational assessment systems, the course could have showed relevance towards the use of tests that are not developed by the classroom

teachers, but from the institutional authorities. Language programme managers should consider working collaboratively with language teachers to enhance their testing resources to make them more efficient, practical and authentic. As an example, participants became aware of misuses and malpractices of assessments in the classroom. The number of students to assess and provide feedback was the most common in the public sector. Participants discussed they have found ways to group their students and promote other assessment practices such as self-assessment and peer- assessments. Combe, et al., (2014) indicated that the use of professional courses might not have a desirable impact on the participants for other reasons such as their own interest and sometimes because they tend to be forced to complete these courses as part of their promotion.

Finally, the last emerging theme was related to impact of the course on the socio-political aspects related to language education in the State of Tamaulipas, Mexico. The findings showed that the participants from the public sector had more difficulties in assessing students' performance adequately. This suggests that assessment is being influenced from different factors such as the number of students, the lower number of hours to empower learning and more interestingly the poor-quality/lack of feedback teachers deliver. It is assumed that the standards of assessment literacy in assessment for learning, alternative assessments and communication of results are the most affected. Participants indicated that the number of students is the main factor that affected students' progress, teachers do not dedicate enough time in the practice of feedback, assessing homework and other extensive activities. These results were also found in studies conducted from (Martinez Rizo & Blanco, 2010)) in which the number of learners in the classroom directly affects the students' performances. They also suggested that the government should react on this matter by decreasing the number of students, as well as more facilities so as learners can obtain better educational practices, especially the language assessment in the public sector. Other research studies have also pointed a lack of preparation from the public sector (Moore, 2017; Roux, 2018).

The number of educational reforms and regulations has impacted teachers' perceptions about the Mexican educational system. Their discomfort has led to develop a lack of interest in professionalism due to the fact they are being forced to be evaluated with national and international standards, but teachers have not received enough support so as to develop their professionalism. Unfortunately, the Federal Government has closed the doors to the National Institute for Educational Evaluation (INEE) creating a perspective that evaluation should be less important. Participants pointed out that the workshop and how Mexico has evolved in terms of evaluation could bring a large benefit to those teachers who are being trained in undergraduates or for professional development towards in-service teachers.

## 7.6 Contributions from these Research Studies

This section has been divided into three parts. The first one discusses the contribution towards the knowledge of language assessment literacy. The second section discuss the contributions towards knowledge and skills for continuing professional development. The last subsection discusses the contributions towards the use of methodological decisions to explore the field of language assessment literacy.

### 7.6.1 Contribution towards Language Assessment Literacy

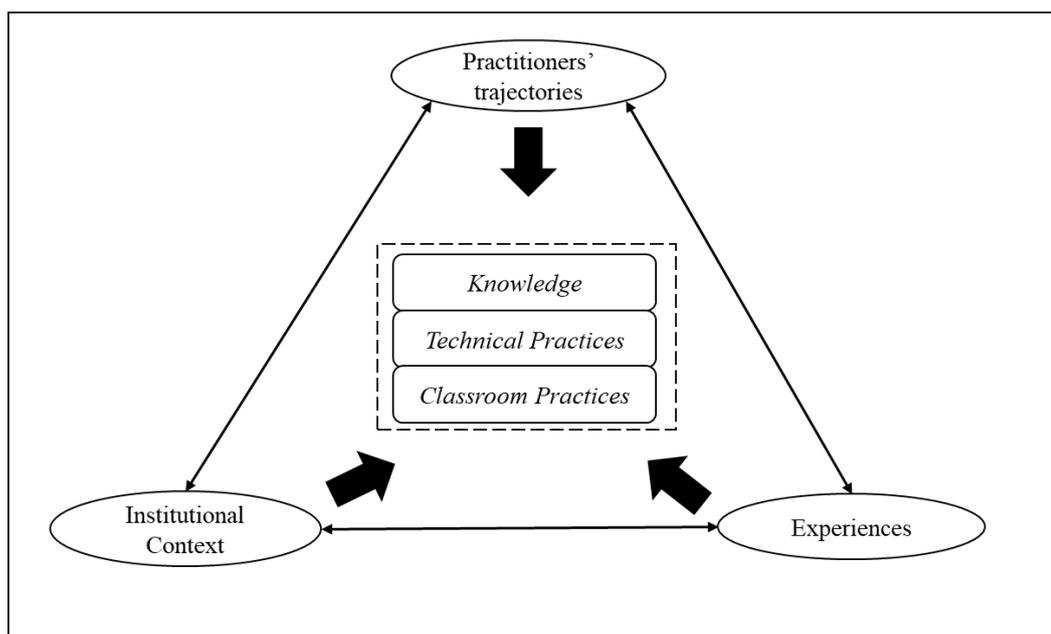
The purpose of this research study was to *examine* the status quo of language assessment literacy in Tamaulipas, to *explore* how EFL teachers self-perceived their language assessment skills and practices (TLASP), to *scrutinise* EFL teachers educational sectors and discover, if any, discrepancies among them, to *identify* any relationships between EFL professional development and language assessment, to *examine* which areas and progress of language assessment EFL teachers require to empower their language assessment literacy, and to *measure* the impact of a contextualised workshop for teachers' language assessment literacy. With a total of five research questions regarding the status quo, educational sectors and how professional development and language assessment literacy interact each other, this research provided clear implications to identify that even some institutions assess differently, the assessment knowledge and skills are fundamental features of educational assessment.

Fulcher's (2012) definition of assessment literacy has included four important aspects that were previously discussed in this study, the knowledge, skills, principles and the context. However, it is the teaching context that plays a crucial aspect for defining assessment. In this research, three educational sectors were examined, the public, the private and the language institute sectors. Similarities and differences were discovered among the participants, these three contexts had underpinned the foundations of assessments in relation to how skilled they perceived and how often they do conduct assessments. It is within the context that assessment literacy can be defined. Assessment literacy is a repertory of competences, skills and abilities that allow students, teachers, language programme managers and other authorities to be informed on how their institution performs (Stiggins, 2002). The context is the key to identify what teachers should really need to become assessment literate, but it is also the feature which differs from other teaching contexts and current needs (Leong, 2014). Exploring three educational sectors or three different contexts under the same umbrella of education in one Mexican State, highlighted that there is a lot of work to be done, but standardising the context could allow teachers interact precisely on what they really need.

The educational context could include aspects that have not been explored yet in different teaching contexts worldwide. This research conducted in the state of Tamaulipas EFL teachers demonstrated that regardless to their different contexts, discrepancies were discovered, but for some teachers these differences might not mean essential for one teaching classroom, but for other might be relevant for their expertise. Language assessment literacy research has expanded worldwide, where a continental research was conducted (Pill & Harding 2013; Tsagari & Voyght, 2014; Berry, Sheehan & Munro, 2017), to examine and analyse teachers' backgrounds to provide one general perspective on how EFL teachers assess their students and what teacher need to maximise language learning proficiency. Other studies have been exploring one country and these researches can provide large benefits to underpin the assessment process and purposes at national level (Plake & Impara, 1991; Lam, 2015; Herrera & Macias 2015), but exploring three educational sectors where the language assessment context varies, it provide an x-ray on how teachers from the same location do to empower assessment literacy. Therefore, this research attempt to explore and examine three contexts and provided a current perspective to understand the purpose of educational assessment in language teaching, the rationale of the use of different assessment practices, how different contexts work, and how could they benefit from each other, and when EFL teachers can identify they have become aware of language assessment literacy.

### **7.6.2 Contribution towards Professional Development**

From these research findings and the materials collected from the workshop, a framework was developed to maximise EFL teachers' professional development in the area of language assessment literacy highlighting three main components knowledge, technical considerations and classroom practices. The main contribution to the field of assessment literacy and professional development is presented in the outer triangle that is taking into consideration teachers' previous experiences as a learner, practitioners' trajectories and the institutional context. Figure 8 illustrates this framework that is entitled as Teachers' Language Assessment Literacy Development Framework (TLALDF). This framework has two dimensions. The language assessment literacy components in the inner side of the triangle and the teachers' backgrounds in the outer side.



**Figure 8 Teachers' Language Assessment Literacy Development Framework (TLALDF)**

The first angle to be discussed is the teachers' experiences. Experiences from being a language learner could allow teachers in building solid foundations of the possible testing resources to be used in the language classroom (Martinez-Rizo, 2012). Previous experiences could lead to empower current perspectives in teaching, learning and assessing the language. Nevertheless, it is important to bear in mind that the educational trends are moving and evolving with the inclusion of technology, self-assessment and the institutional needs.

Then, the second angle of this triangle deals with important to highlight the teachers' academic and professional trajectories. It is important to highlight that for the purpose of this research, practitioners' trajectories focus on the teachers' academic and professional pathways. Teachers' backgrounds are essential to identify and observe to which extent teachers are assessment literate and how their academic and professional trajectories have modelled the methods of developing and conducting language assessments (Moore, 2017). In the case of Tamaulipas and in other regions of Mexico, EFL classrooms can be led by multi-disciplinary teachers who were not involved in teacher training during their undergraduate or they have a lack of these continuing professional development opportunities.

Lastly, the last angle of this triangle refers to the institutional context, this is where the operational assessment system teachers have access in their classrooms. Literature has suggested that sometimes teachers do not develop their own tests, but the importance to know about the process of developing, as well as taking into consideration the practices of how assessment takes place is important to highlight to understand the benefits of educational assessment (Stiggins, 2002). In addition to TALD framework, I also proposed self-evaluation descriptors to empower

language assessment literacy in the language classrooms. These descriptors were divided into the three categories of the framework: knowledge, technical and classroom practices. These descriptors were also developed based on the findings from this study in relation to the outer triangle that was mentioned abovementioned (see Figure 8). This can be used not only for teachers to locate their assessment skills and practices, but also used by the language programme managers to identify the essential aspects teachers need regarding assessment and testing for different institutions and contexts.

In the inner dimensions includes the three components of the language assessment literacy: knowledge, technical considerations and classroom practices which were presented (see Figure 2). Each of these three components has different language assessment literacy standards (AFT, 1990; Giraldo, 2018). As discussed in Chapter 2, knowledge of assessments stands for three standards; applied linguistics, educational theories and institutional policies. The technical practices refer to the standards of testing principles, assessment processes, assessment purposes and educational management. The last component is classroom assessments that deals with the standard of assessment for learning, communication of results, language assessment and assessment as educational supports. These components were chosen based on the models of DeLuca (2013) and Giraldo (2018) oriented towards the assessment of language in the classroom.

As discussed in the previous section, this research project proposes the Teachers' Language Assessment Literacy Development Framework (TLALDF) and the self-evaluation descriptors for teachers' assessment skilfulness and practices in the classroom (see Appendix M). The descriptors were developed based on the framework that guided this research project. These distractors could open the possibility to be tested and examined with different educational sectors and teaching contexts from primary, secondary and tertiary levels. The findings of this research project have clear implications for local policy makers in the Mexican State of Tamaulipas. The Ministry of Education in Tamaulipas (SET) could evaluate the findings of this research to support the educational sectors in this State. In addition, the results can also complement the recommendations from the OECD in relation to the public sector. The findings have demonstrated that participants from the public sector tend to work extra hours for planning, developing materials, taking extra continuing professional development but they do not prioritise the feature of assessment. However, the extra amount of work from teachers especially those who work in high school and several higher education institutions, can affect the performance of the students when assessments are not accurately administered. The large number of students in the classroom, teachers' working hours and the lack of training can lead to obtain poor students' results in classrooms and national assessments.

### **7.6.3 Contributions towards Methodological Decisions**

This research study followed two different research methodologies to answer the five research questions. The exploratory research used a mixed-method approaches to analyse the quantitative data from a survey TALDI and qualitative data from semi-structured interviews. The experimental research used a concurrent mixed-methods which use quantitative and qualitative data: the materials from the workshop (e.g., pre- and post- measures, reflective and practical tasks and a workshop evaluation sheet) and a focus group interview that was able to corroborate and to expand the findings from the workshop materials. Use of these two methodologies made possible to support different research aspects such as the use of language tests among the three educational sectors and the inclusion of continuing professional development to explore the status quo of a geographical region. By conducting research from this scope, the concurrent and mixed-methods of data analysis can provide a lot of benefits to examine different aspects from the language assessment literacy field.

Research in LAL has used a diversity of research methods which mixed-methods have been the most appropriate (Mertler & Campbell, 2005; Malone, 2013; Berry, Sheehan & Munro; 2017); however, using concurrent mixed-methods, as an additional ingredient to the research, has the opportunities to corroborate and support findings as the way this research study applied.

## **7.7 Limitations of the Study**

This section provides a brief analysis of several limitations of this current study. The first limitation is the scope of the study which takes into consideration one geographical zone, Tamaulipas, three educational sectors and theoretical aspects that are considered fundamental in language assessment literacy. It is important to highlight that even though the research questions were answered based on the data collected and analysis of the findings, by limiting the scope of this study could provide detailed and contextualised findings.

Other limitation is that this survey entitled TALDI, which has been adapted for the purpose of this research, is a self-perception survey; in order words, this instrument does not examine or tests participants' knowledge, but it has a high reliability index from previous studies (Zhang and Burry-Stock 2003), as well as, for this research project. This quantitative instrument was piloted, and statistical tests were conducted to measure its reliability index through Cronpha Alpha. This is important to highlight due to the fact that participants can lie or alter their perceptions.

Although some literature has suggested to conduct classroom observations to collect evidence on the current situation in relation to assessment, these were unable to be collected

due to institutional restrictions from the Mexican educational sectors. Conducting classroom observations can be a great source of analysing important information. However, I believe that the contextualised training course was an opportunity to explore each sector differently and see what the current assessment practices teachers were conducting in their classrooms during a teaching term.

Another important aspect that needs to be mentioned is the relationship between the trainer of the workshop and the researcher. Since I was responsible for two activities, I tried to be as much objective with the progress of the participants, as well as, measuring the impact of the course. Consequently, the focus group was conducted 14 weeks after the intervention to identify any positive or negative changes in a group of participants. I strongly believe that by conducting this focus group, I was able to provide more support to teachers and make this slightly less biased.

## **7.8 Opportunities for Future Research**

This study can also have the opportunity to be replicated in other Mexican States with the assistant of the quantitative instrument Teachers' Assessment Literacy Development Inventory (TALDI) which includes essential aspects not only to identify teachers' assessment skills and practices, but also to be able to explore teachers' professional development needs and to amend the contextualised course to develop a National course for Mexican teachers of English. A new incorporation to the features of language assessment could be the use of technology to revise and administer language tests

This study can also be the engine to explore the educational sectors (e.g., public, private and language institutes) in depth. Taking in consideration the geographical zone and regions in Tamaulipas or in other Mexican States to explore other regions such as the rural and urban regions and to what extent teachers and language programme manager require to maximise teaching and learning efficiency. Also, since this study included teachers and language programme managers, students can also be considered to triangulate and to reinforce teachers' and students' needs. A proper triangulation of these three stakeholders would bring large benefits to the field of language testing and assessment for EFL classrooms not only in the State of Tamaulipas, but at national scale. Lastly, this study can further be developed by including classroom observations to identify whether the contextualised course had a direct impact on the students. Classroom observations have been suggested and applied by different researchers in the field of language testing and assessment.

The teachers' assessment skills and practices varied among educational sectors, but this could force language programme managers to work collaboratively with their teachers to improve the

quality of assessment by aligning tests and other assessment types into the curriculum and objectives of the course. Language assessment literacy has opened a new panorama for language educators to evaluate and reflect on their own practice with the intention to identify what teachers current needs are. It is crucial that teacher preparation courses are well appointed with the foundations but at the same time they need to have a clear purpose and objective to cover those specific needs from one region or state so as teachers have the opportunity to expand their views to understand and identify what the field (language assessment literacy) requires to become assessment literate.



# Appendix A Participants' Consent Form



## CONSENT FORM (F2F: 25427-V01)

Study title: LANGUAGE ASSESSMENT LITERACY DEVELOPMENT AMONG MEXICAN TEACHERS OF ENGLISH IN HIGHER EDUCATION: EXPERIENCES, PRACTICES AND KNOWLEDGE

Researcher name: RICARDO DE LA GARZA CANO

Staff/Student number: 26464551

ERGO reference number: 25427

*Please tick the boxes if you agree with each statement:*

I have read and understood the information sheet ( Participant Information Sheet (F2F: 25427 -01)) and have had the opportunity to ask questions about the study

I agree to take part in this research project and agree for my data to be used for the purpose of this study .

I understand my participation is voluntary and I may withdraw at any time without my legal rights being affected.

### *Data Protection*

*I understand that information collected about me during my participation in this study will be stored on a password protected computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.*

Name of participant (print name).....

Signature of participant.....

Date.....

## Appendix B Participant Information Sheet

### PARTICIPANT INFORMATION SHEET (F2F: 25427-01)

Study Title: **Language Assessment Literacy Development among Mexican Teachers of English in Higher Education: Experiences, Practices and Knowledge**

Researcher: RICARDO DE LA GARZA CANO      Ethics number: 25427

Please read this information carefully before deciding to take part in this research. If you are happy to participate, you will be asked to sign a consent form.

#### **What is the research about?**

This research is necessary in order to meet the requirements to successfully achieve the PhD Programme Degree in Modern Languages from the University of Southampton. The purpose of this study is to analyse the status of language testing and classroom based language assessment in Higher Education (HE) institutions in Mexico. This investigation will collect evidences from three different stakeholders' perspectives (Head of departments, language instructors and language learners) in several Mexican Universities.

#### **Why have I been chosen?**

You have been chosen to participate in this research study since you meet all the requirements that this investigation involves. In order words, you should be working in a Mexican HE institution, either public or private. More importantly, you MUST be a head of department, language instructor or language learner, primarily in the English language.

#### **What will happen to me if I take part?**

This study is divided into three phases. The first phase takes part as a semi-structured interview to collect different perspectives and beliefs related to language testing and classroom-based language assessment (not mandatory). Only for the first phases, head of departments and language learners are allowed to participate. In addition to this, you will be answering a questionnaire that aims at collecting several assessment practices and knowledge related to languages testing and assessment (mandatory). The second phase is to participate in a Language Testing and assessment on-line or face –to-face course, in which several tasks need to be completed and you will receive valuable feedback. This course is totally free and will be via blackboard or face-to-face. The third phase is to collect evidences as to whether the course made an impact on the participants. If you accept to participate and complete all the tasks in the course, you will be interviewed (semi-structured or focus group) to scrutinise any impact of the course for the language teaching and assessing practices.

**Are there any benefits in my taking part?**

If you choose to participate in this investigation, you will be invited to take a free course to improve your language testing and classroom based language assessment practices.

**Are there any risks involved?**

There are no high risks for this study.

**Will my participation be confidential?**

All personal information such as participants' name, gender, workplace, and personal responses will be confidential and will be stored in a computer from the University of Southampton with a password that only the researcher will have access to.

**What happens if I change my mind?**

If you have participated in any of the three phases and you suspend, you need to notify the researcher (Mr. Ricardo de la Garza Cano) via email ([rdlg1g14@soton.ac.uk](mailto:rdlg1g14@soton.ac.uk)). If you have been scheduled to be in an interview, please send an email 72 hours before the day of the task.

**What happens if something goes wrong?**

In the unlikely case of concern or complaint, you should provide a named independent contact with phone number and email address to the Chair of the Faculty Ethics Committee, Prof. Denis McManus ([D.Mcmanus@soton.ac.uk](mailto:D.Mcmanus@soton.ac.uk)).

**Where can I get more information?**

If you have any questions regarding the structure of this study, data collection or confidential information, you can email the main supervisor of this study, Dr Ying Zheng ([ying.zheng@soton.ac.uk](mailto:ying.zheng@soton.ac.uk)) or the researcher, Mr. Ricardo de la Garza Cano ([rdlg1g14@soton.ac.uk](mailto:rdlg1g14@soton.ac.uk))

**Thank you very much!**

## Appendix C Example of an Approval Participation

Por medio de este conducto solicito su apovo para realizar un estudio piloto y final en en Ciudad Victoria, Tamaulipas, Mexico. Este proyecto se titula **"Literacidad en la evaluation de la lengua Jnglesa entre Jos profesores mexicanos en educaci6n superior: experiencias, practicas y conocimiento"**.

Este analizara las perceptivas, practicas y conocimientos que los profesores de lengua Inglesa en la educaci6n superior en Mexico tienen y deben de aplicar para poder llevar una evaluaci6n justa, concreta y eficaz.

Para dicho proyecto, se realizara la aplicaci6n de encuestas a la totalidad de sus profesores. De igual manera, se realizaran un maximimo de 5 entrevistas, una al coordinador(a) del programa de lenguas, a dos profesores que se encuentren activamente trabajando en dicho programa y a dos de sus estudiantes. Estas encuestas seran analizadas en la ciudad de Southampton y por el investigador Mtro. Ricardo de la Garza Cano, estudiante de Doctorado de la Universidad de Southampton en el Reino Unido, los resultados seran totalmente confidenciales ya que no se mencionaran informaci6n ni de las participantes, instituciones o respuestas por ningun medio.

Sin embargo, si la instituci6n que acepte dicha participaci6n a este proyecto quiera obtener un reporte escrito del desarrollo profesional de los profesores, este se les dara una vez terminada la etapa de analisis de datos (Marzo – Agosto 2018).

Me despido de usted agradeciendole su apoyo brindado.

  
Mtro. Ricardo de la Garza Cano



English Translation of the Letter of Approval\*

(Director's or Language Programme Manager's name)

Through this channel I request your support to conduct a study Pilot and final in the (name of the institution) in Ciudad Victoria, Tamaulipas, Mexico. This project is entitled "Literacy in the evaluation of the English language among Mexican teachers in higher education: experiences, practices and knowledge".

This will analyse the perceptions, practices and knowledge that English language teachers in higher education in Mexico have and they must apply in order to carry out a fair, concrete and effective evaluation. For this project, the application of surveys to the all your teachers. Similarly, 5 interviews will be conducted, one to the coordinator of the language program, to two teachers who are actively working on that program and two of its students.

These surveys will be analysed in the city of Southampton and by researcher Mtro. Ricardo de la Garza Cano, student of Doctorate from the University of Southampton in the United Kingdom, results will be totally confidential since they were not mentioned information from participants, institutions or responses by any means, medium.

However, if the institution that accepts such participation to this project wants to obtain a written report of the professional development of the teachers, this will be given once the data analysis stage is over (March - August 2018).

I say goodbye to you thanking you for your support.

---

Mtro. Ricardo de la Garza Cano

***\*Note: This translation was developed by the researcher.***



## Appendix D Survey (TALDI)

### SECTION 1. Participants' Backgrounds

INSTRUCTIONS: Cross  the most appropriate answer based on your profile.

1. What type of educational sector is the institution where you work?

- Public     Private     Language Institute

2. What is your gender?

- Male     Female

3. How old are you?

- Under 25     25-29     30-39     40-49     50-59     60+

4. What is the highest level of formal education that you have completed?

- Bachelor's Degree     Master's Degree     Doctorate's Degree  
 Other, Please specify \_\_\_\_\_

5. What is the field of your educational degree?

- |  |  |
|--|--|
| <input type="checkbox"/> Applied Linguistics       | <input type="checkbox"/> Pedagogy          |
| <input type="checkbox"/> English Language Teaching | <input type="checkbox"/> Foreign Languages |
| <input type="checkbox"/> Social Sciences           | <input type="checkbox"/> Other             |
| <input type="checkbox"/> Basic Education           | Please specify _____                       |

6. What is your current employment position?

- Language Teacher     Manager     Administrative position  
 Other, Please specify \_\_\_\_\_

7. How many years of teaching experience do you have?

- Less than a year     1 - 2     3 - 5     6 - 10     11 - 15     16 - 20     More than 20

8. What is your employment status?

- Full-time     Part-time

9. What is the average number of students you have in a language classroom?

- 1 - 10     11 - 15     16 - 20     25 - 30     31 - 40     40+

10. How many hours do you teach per week in front of a classroom?

- 1 - 5     6 - 10     11 - 15     16 - 20     21 - 29     30 - 40

11. Which language level(s) based on the Common European Framework (CEF) are you currently teaching? [More than one option is possible]

- A1     A2     B1     B2     C1     C2

### SECTION 2. Participants' Experiences and Educational Context

Answer these questions based on your assessment practice in the institution where you are working.

12. How do you assess your language students?

.....

13. Do you design the tests that are administered in your classroom? If not, who develops the tests?

.....

14. Apart from language tests, what else do you use to assess your students?

.....

15. How do you choose the methods of assessment?

.....

16. How do you realise that the methods for testing and assessing your students are the best?

.....

17. Are there any code of practices for development, administration, marking or communication of assessments in your institution?

.....

18. Have you identify any problems with the use of assessments in your institution?

.....

Survey (TALDI)

**SECTION 3. Professional Development in Language Testing**

19. Are you interested in language testing and classroom-based language assessment practices?

Yes  No

20. Which of the following statements best describe your perception of your overall knowledge, understanding and application of language testing?

Very unprepared  Somewhat unprepared  Somewhat prepared  Very prepared

21. Which of the following statements best describe your perception of your overall knowledge, understanding and application of language assessment?

Very unprepared  Somewhat unprepared  Somewhat prepared  Very prepared

22. During the last 18 months, have you attended any workshop(s), course(s), university module(s) or conference(s) fully related to language testing and assessment?

Yes  No

22a. If you have crossed (Yes) for question (21), What is the name of the organisation, association or institution you attended?

23. From the following list of language testing and assessment related topics, which ones do you consider very essential or not essential for a suitable teacher-training course you would attend?

	Not at all essential	Little essential	Somewhat essential	Essential	Very essential
History of language testing					
Identify and select contents for testing					
Suitable methods for assessing language					
Development of language tests					
Writing test specifications					
Effectiveness of language tests (item analysis)					
Types and purposes of items, tasks and tests					
Administration of language tests					
Interpreting test-takers' scores and performances					
Assess receptive skills (reading and listening)					
Assess productive skills (writing and speaking)					
Assess integrated skills					
Assess language components (vocabulary, grammar and pronunciation)					
Assess students with special learning needs					
Assess students through portfolios					
Assess students through self-assessment					
Assess students through peer assessment					
Language tests and statistical approaches					
Language assessment principles					
Using analytic scales for productive skills					
Using holistic scales for productive skills					
Preparing learners for standardised tests (FCE, etc.)					
Ethical considerations in testing					
Reporting students' performances (feedback)					
Understanding the Common European Framework bands					
Interpreting test-takers' scores to amend teaching practice					

**SECTION 4. Continuing Professional Development**

24. Are you interested in your professional development?

Yes  No

25. Is professional development compulsory in the institution where you work?

Yes  No

26. Does your institution provide support (e.g., financial, mentoring, etc.) to improve your professional development?

Yes  No

26a. If you have crossed the box (Yes) for question (19), How does your institution support your professional development?

.....

.....

.....

27. Which of the following English language certifications have you achieved? [More than one option is possible]

KET  PET  FCE  CAE  CPE

TOEFL  IELTS  Pearson Academic  CENNI

Other, please specify \_\_\_\_\_

28. Which of the following English language teaching certifications have you achieved? [More than one option is possible]

TKT  CELTA  COTE/ICELT  DOTE/DELTA  TESOL/TEFL Certificate

Other, please specify \_\_\_\_\_

29. From the following list of professional development features, categorise from very important to not important for your teaching practice:

	Not at all important	Unimportant	Neither important nor unimportant	Important	Very important
1. Planning lessons	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Knowledge of the subject	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Materials and resources for language teaching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Language testing and assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Approaches in language teaching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Understanding language learners	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Technology in language education	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Educational policies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Classroom management	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Autonomous learning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Other, please specify _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

30. During the last 18 months, what kind of activities from the following list have you participated in to improve your professional development, and to what extent have these activities had an impact on your teaching practice?

	Participation (A)		Impact (B)				
	Yes	No	No impact	Small impact	Moderate Impact	Large impact	Significant impact
1. Courses/ workshops	<input type="checkbox"/>						
2. Conferences	<input type="checkbox"/>						
3. Qualification programmes (B.A., etc.)	<input type="checkbox"/>						
4. Observations	<input type="checkbox"/>						
5. Online course/ webinars	<input type="checkbox"/>						
6. Reading books, research articles	<input type="checkbox"/>						
7. Mentoring, peer-observations	<input type="checkbox"/>						
8. Other, please specify _____	<input type="checkbox"/>						

Survey (TALDI)

31. From the following list of National and International organisations, research groups and companies, which ones have you attended as part of your professional development and to what extent have they had an impact on your teaching practices?

	Participation (A)		Impact (B)				
	Yes	No	No impact	Small impact	Moderate Impact	Large impact	Significant impact
MEXTESOL	<input type="checkbox"/>						
British Council	<input type="checkbox"/>						
Red Nova	<input type="checkbox"/>						
SEP and CENNI	<input type="checkbox"/>						
Cambridge University Press	<input type="checkbox"/>						
Private Consultants	<input type="checkbox"/>						
RECALE	<input type="checkbox"/>						
Other, please specify _____	<input type="checkbox"/>						

32. From the following list of learning activities, which ones best fit your learning style? [More than one option is possible]

<input type="checkbox"/> Presentations (PPTs)	<input type="checkbox"/> Online resources	<input type="checkbox"/> Group tasks	<input type="checkbox"/> Books
<input type="checkbox"/> Videos	<input type="checkbox"/> Individual tasks	<input type="checkbox"/> Portfolios	<input type="checkbox"/> Research articles
<input type="checkbox"/> Reading groups	<input type="checkbox"/> Discussion groups	<input type="checkbox"/> Note taking	<input type="checkbox"/> Other(s)

Please specify \_\_\_\_\_

33. What kind of learning environment best describes your studying habits?

Face-to-face   
  E-learning   
  Blended learning   
  Other,

Please specify \_\_\_\_\_

34. Would you like to participate in a free training course for language testing and assessment?

Yes   
  No

34a. If you have crossed the box (Yes) for question (29), how many hours overall would you be able to spend in a course for language testing and assessment practices?

0 – 10   
  11 – 19   
  20 – 29   
  30 – 39   
  +40

34b. If you have crossed the box (Yes) for question (29), for a professional development course, how many hours would you spend per daily session?

1   
  2   
  2 - 4   
  4 - 6   
  +6

34c. If you have crossed the box (No) for question (29), what would be the reasons for not participating in a language testing and assessment course?

Lack of interest   
  Lack of time   
  I don't find this useful   
  Other, please specify \_\_\_\_\_

**SECTION 5. Practitioners' Assessment Skills and Practices**

**INSTRUCTIONS:** Cross  the most appropriate answer based on your classroom language assessment practices. In this section, there are two main questions. The first one deals with the frequency of such activities, **how often do you do the following assessment practices?** (*Never to very often*). The second one deals with the level of knowledge and skills to perform such activity, **how skilled are you in doing these activities?** (*Not at all skilled to very skilled*).

	How often? (A)					How skilled? (B)				
	Never	Almost never	Sometimes	Often	Very often	Not at all skilled	A little skilled	Somewhat skilled	Skilled	Very skilled
<b>Applied Linguistics</b>										
35. Clarify language use and application into the classroom	<input type="checkbox"/>									
36. Application of methods for learning, teaching and assessing (memorisation, drills, task-based learning)	<input type="checkbox"/>									
37. Theoretical knowledge and application of learning trends (behaviourism, constructivism and socio-cultural)	<input type="checkbox"/>									
38. Second language theories affective approaches (motivation, preparation, learning efficiency methods)	<input type="checkbox"/>									
<b>Educational Assessment</b>										
39. Awareness of assessment for/of/as learning	<input type="checkbox"/>									
40. Understand the underpinning of content and purpose of assessments	<input type="checkbox"/>									
41. Plan and implement materials for assessing language in relation to the learning context	<input type="checkbox"/>									
42. Enhance and maximise learning through assessment practices	<input type="checkbox"/>									
<b>Institutional context</b>										
43. Understand of the socio-political aspects of assessment towards national policies	<input type="checkbox"/>									
44. Apply national and institutional standards into the language classroom	<input type="checkbox"/>									
45. Distinguish assessment methods that benefit individuals, institutions and educational sectors	<input type="checkbox"/>									
46. Participate in development of tests and assessments for classrooms purposes	<input type="checkbox"/>									
<b>Language Assessment Principles</b>										

Survey (TALDI)

47. Identify the properties and usefulness of a language test by applying the testing principles	<input type="checkbox"/>									
48. Demonstrate awareness of the relevance of the principles at the moment of developing language test.	<input type="checkbox"/>									
49. Classify language tests based on their usefulness and plication towards learning.	<input type="checkbox"/>									
50. Judge the usefulness of language tests towards teaching and learning practices.	<input type="checkbox"/>									
51. Apply knowledge to highlight irregularities and efficiency in language tests.	<input type="checkbox"/>									
<b>Assessment Purposes</b>										
52. Understand the differences of assessment purposes	<input type="checkbox"/>									
53. Apply accurate assessment method such as diagnostic, formative or summative	<input type="checkbox"/>									
54. Choosing the purpose of tests based on curricula and learning needs.	<input type="checkbox"/>									
55. Contrast and examine differences among types of assessments	<input type="checkbox"/>									
56. Examine the different assessment purposes to match the teaching and learning objectives	<input type="checkbox"/>									
<b>Assessment Processes</b>										
57. Understand different properties and usages of language tests.	<input type="checkbox"/>									
58. Differentiate and choose appropriate criteria.	<input type="checkbox"/>									
59. Underpin the purpose with the method for assessing a particular language skill or linguistic component.	<input type="checkbox"/>									
60. Awareness and application of different types of items and tasks to be included in language assessments.	<input type="checkbox"/>									
61. Infer and make educational judgments about the qualities of test.	<input type="checkbox"/>									
<b>Educational Measurement</b>										
62. Awareness of methods for evaluation of items an tasks	<input type="checkbox"/>									
63. Identify the level of difficulty of items and tasks	<input type="checkbox"/>									
64. Differentiate between higher and lower scores to discriminate items	<input type="checkbox"/>									
65. Conduct analysis to measure the quality of items and tests	<input type="checkbox"/>									
66. Distinguish functional and non-functional items and tasks	<input type="checkbox"/>									
<b>Language Assessment</b>										
67. Recognise the properties of assessment language receptive and productive skills.	<input type="checkbox"/>									
68. Use and adapt criteria for marking productive skills (reading and listening)	<input type="checkbox"/>									
69. Choose types of responses based on the purpose of assessments.	<input type="checkbox"/>									
70. Justify the use of assessment methods for linguistic features (e.g., grammar, vocabulary, etc.)	<input type="checkbox"/>									

71. Manage to assess linguistic and content properties based on the assessment methods	<input type="checkbox"/>									
72. Recognise the properties of assessment language receptive and productive skills.	<input type="checkbox"/>									
73. Integrate skills in language tests and assessments	<input type="checkbox"/>									
<b>Assessment for Learning</b>										
74. Development of alternative assessments to promote learning	<input type="checkbox"/>									
75. Distinguish properties among alternatives (performance assessments) and standardised tests to prioritise learning	<input type="checkbox"/>									
76. Develop awareness of the advantages of portfolios and other assessment tools to collect learning evidence	<input type="checkbox"/>									
77. Prepare and predict students' performances based on alternatives for assessment, excluding language tests	<input type="checkbox"/>									
78. Align alternative assessments to a specific purpose of assessment (e.g., diagnostic, formative and summative)	<input type="checkbox"/>									
<b>Communication of Assessments and Results</b>										
79. Inform students about the assessment practices and methods for marking before and during the course	<input type="checkbox"/>									
80. Communicate test-scores and other assessments marks to students	<input type="checkbox"/>									
81. Inform students about their progress (feedback) to promote learning	<input type="checkbox"/>									
82. Measure the impact of the tests towards students' learning trajectories	<input type="checkbox"/>									
83. Inform institutional authorities about students' lack of progress	<input type="checkbox"/>									
<b>Assessment as an Educational Support</b>										
84. Compare test scores and other assessment to amend teaching, learning and assessing methods in the classroom	<input type="checkbox"/>									
85. Interpolate students' progress and teaching assessing methods to see efficiency in learning	<input type="checkbox"/>									
86. Examine and exchange students' performances to improve the quality of tests and other sources	<input type="checkbox"/>									
87. Recommend and suggest further development of assessment methods based on students' performances	<input type="checkbox"/>									
88. Evaluate the impact of tests for learning, teaching and assessing practices	<input type="checkbox"/>									

## **Appendix E Interview Guidelines**

### **A) Training in Language Testing and Assessment**

- 1) Do you consider professional development important for your career?
- 2) Have you taken a training course in LTA in the last 18 months?
  - a. If yes, which organisation
  - b. If not, why (reasons)
- 3) Do you think a training in LTA is important for your teaching practice?
- 4) Have you taken any teaching qualifications in which assessment is included?
- 5) In your opinion, why haven't you taken a specialised module in LTA?

### **B) Language Tests: Importance and frequencies**

- 1) What is your perception towards language tests? Are they important?
- 2) How often do you assess your students with language tests?
- 3) Do you use tests for formative or summative purposes?
- 4) How many tests do you consider your students should take in a language course?

### **C) Language test development and Assessment for language Skills**

- 1) Do you develop the language tests?
  - a) If yes, how do you do them?
  - b) If not, who does this?
- 2) What language skills do you assess?
- 3) How do you assess your students with tests?

### **D) Teachers' Assessment Knowledge and Technical Skills**

- 1) What do you consider it is important for teachers to know about LTA?
- 2) From a scale 1 to 5, how skilful you consider in LTA?
- 3) How do you know if the tests are accurate for your students?
  - a. Do you conduct item analysis or any other statistical procedure?
- 4) How well you know the principles for testing?
- 5) How well do you assess your students in relation to receptive & productive skills?

### **E) Teachers' Classroom Practices**

- 1) Apart from language tests, what else do you use for assessing your students?
- 2) How well do you distribute your formative and summative assessments?

### **F) Teachers' Assessment Experiences**

- 1) How do you know which methods for assessment suits your students?
- 2) In your opinion, how do you know if the tests and assessment methods are appropriate for your course which you are teaching?
- 3) In your opinion, what is a bad and a good assessment experience?

### **G) Specific problems in your educational contexts**

- 1) Have you identify any problems related to your institution in relation to LTA?  
Are there any other aspects you would like to share with me about the use and practices of LTA in your institution?

-----THIS IS THE END OF THE INTERVIEW-----

## Appendix F Example of an Interviewee (LI-T4)

**Interviewer:** This is LI-T4, she is 30 years old she is doing an MA in educational innovation. She's an English teacher and she's got 6 years of teaching experience. Hello LI-T4, how are you?

**LI-T4:** Fine thank you and you?

**Interviewer:** I'm very well thanks. This interview collects information about your experience with regards to language testing and assessment. Divided into seven parts. And let's start with the importance of language assessment. So to begin with what do you think about assessment?

**LI-T4:** I think it's very important because we need to let students know about their progress about their work. They need to see that we are really paying attention to their needs and we are covering that aspect in which everything they do is important.

**Interviewer:** All right. Thank you. Are you responsible for testing and assessment in your classroom?

**LI-T4:** I have to apply that test but I don't work in designing them at the moment. I have experience doing that but not in this particular moment. In this institution where I work they have the exams that they designed them in advance.

**Interviewer:** Is assessment an aspect for your teaching practice? Is assessment important or less important and why?

**LI-T4:** It is important I always try to give feedback to my students and always try to make them be aware of their knowledge and their abilities. I also try to make them work with their classmates because I think it's important to help each other and to see how they can improve their work.

**Interviewer:** What are the characteristics of an accurate assessment in classroom?

**LI-T4:** Well I think that the teacher needs to be very clear since the beginning on what he or she is going to observe and to analyse and to take into account and also the student needs to understand every process of assessment. And it's also important that in case there is something that is not clear the student feels confident to ask about it.

**Interviewer:** How much are you involved in assessments?

**LI-T4:** Well I see I have been told by my coordinator to use rubrics I have been using some of them and I'm also working on my own. I try to give my students my conclusions in every important task they perform. And well basically I tried to do at least one of those important activities per unit so they can have enough time to see what they did and what can be changed or modify in practice.

**Interviewer:** What are the assessment procedures for students? And which ones do you take into consideration?

**LI-T4:** Well in general they have to work in class they have to solve some exercises they have to write a lot. Some of the levels I'm teaching now they are writing all the time some is speaking all the time. And the thing is that I try to provide them with different activities to evaluate their work and that test is very important for them but it's not the only thing they do they are working 50/50 with portfolio also, they are used to it. Some people don't like it as students they

#### Example of an Interviewee (LI-T4)

are not used to it but they are getting into it little by little because they see the importance of having more evidences of their knowledge than just a test.

**Interviewer:** Who said what and how teachers need to assess their students in your institution?

**LI-T4:** Well they have evaluation that takes place every three units is in the calendar that this gives at the beginning of the course and they already know I mean the students already know that they know the times where they are going to be evaluated but the rest of the evaluation is performed, Sometimes in academies we determine activity for students. We plan some activities together but there are always two or three activities that are set just by the teacher in the classroom and it depends a lot on the group. I mean the level of the group and the teacher-students interaction at the moment. So he or she or in my case I decide which activities are the ones I'm going to use for assessment.

**Interviewer:** Do you believe that student scores represent the quality of your teaching practice.

**LI-T4** Yes I think that there is a correspondence between my work and their performance because we give each other support I mean I am helping them in the process of learning but they are also helping me in the process of teaching because they helped me improve they make me notice what is good in my practice what I can improve when I can change or what I need to modify somehow and they assess my work even if they don't notice it. They are doing the same thing.

**Interviewer:** How much are you involved in developing language test instruments or materials to track students progress.

**LI-T4:** Well I design some materials as I mentioned before at this moment I'm not designing the official test but I design quizzes and I design tasks for my students and I also adapt and adopt other material. So I'm always trying to find something new to -to show them something new to give them to be in the process of learning. But while the official one is not in it's not something I have to do but the rest of the assessment or the different parts of the evolution that they take is where I can do something about it and I can design or modify my materials.

**Interviewer:** What are the elements you take into consideration for developing language stress instruments and testing materials?

**LI-T4:** Well I always think that the instructions are really important for students when I am designing a task. I also take into account that they have to have something they know, I mean from before from their background and it needs to be challenging at the same time. If it's not challenging, I consider that it's not valid enough because they are just using very -how can I say- like they are solving the task in a very simple way. So it needs to be challenging they really need to think about it they really need to use their logic and their skills to solve the task. So I think it has to be challenging but it also has to be connected to their learning at the moment. And the previous one I cannot ask my students to answer something they have no idea about it.

**Interviewer:** How are language tests administrated to your students in your institution?

**LI-T4:** Well the tests take place every three units and they consist in two parts. The first part is this speaking test, it includes with some cards and instructions that the coordinator gives us, we receive them material the same day we apply the exam. Just a few minutes before. So some teachers sometimes have no idea what counts in the evaluation officially, we have the idea because we know they can't. But I mean we don't know the task in advance. So when we present to the students to work with it we are almost at the same point in which we are discovering what the material includes And then they have the written test. In my institution, as the classes last for two hours it seemed possible to have both test the same day. So it's one day for the speaking the

next for the written and the written is divided into four sections. The first one is use of vocabulary and grammar which is called use of language. The second is reading and they have to listen and finally they have the writing in which they have to answer to certain requirements.

**Interviewer:** How do you make sure that the language test instruments and that you have that you produce. I mean just use requirements for learning.?

**LI-T4:** Well there was a time in which my coordinator and some people were designing the exams and I remember they asked us to verify that the quantity of the exam has really real connection with the material we are using in class. But we also tried to at the moment we try to let them know that they also need to modify some things. I mean that the exam is not just copying what they learned from the book or what they saw in the book but basically the topics I read relate to them very connected.

**Interviewer:** what do you do in order to validate language test instruments in your classroom?

**LI-T4:** Well my students always receive feedback about it in my practice. Well I have seen that they really take advantage of the test to see their progress but also I try to make them aware of what they can change for next time I mean for the next exam. So I think that's a validation of the test comes from confirming that it is evaluating the aspects they saw in class it is evaluating what they are supposed to know what the level. They know that there is a standard level they need to be in and they are aware they are very conscious of themselves and their progress and they know if they are achieving their goals or if they are not.

**Interviewer:** Do you analyse your students' results to prepare new materials or instruments?

**LI-T4:** Yes, I will take the results in order to see which are the areas in which they are failing or they are presenting weakness. So what I do is that I try to reinforce that before getting to the next topic because otherwise I think that they're not getting the previews they're not getting the new one. So there's going to be a gap in between. And what I do is that I always try to reinforce before and after that test. So maybe for the number they got it's not going to change but that process is going to improve and they are going to get a better result next time.

**Interviewer:** Do you analyse students results to amend your teaching, learning and assessing methodologies?

**LI-T4:** Yes one of the aspects I am always worried about is grammar. So I always observe which way of presenting grammar is the best for them. I know that they I am evaluating other areas but this is the one in which students have more problems most of the time. So what I do is that I see the ways in which I can freeze grammar in certain tests. If they didn't get a good result then I know that my presentation wasn't as good as I thought at the beginning. I mean it's not all my responsibility but when I observe that it's like in general terms like 80 per cent of the students have trouble with something then either it's me or is the book or both of us but I need to do something about it.

**Interviewer:** What kind of assessment tools do you use in the classroom? **PUB-T2:** Well we have the rubrics, we have quizzes we have they have to write reflections on their progress so that is the self assessment. They also have to check each other's work from time to time so that is the peer assessment and that is something that I consider very useful and in my experience even when I don't ask them to help each other they do. And there's always a one

#### Example of an Interviewee (LI-T4)

or two of them in each team that are able to help their classmates and they really want to do it. So it's like it comes in a natural way. They become friends more than classmates and that helps each other to do that.

**Interviewer:** What kind of feedback do you see in the classroom?

**LI-T4:** Well I try to provide different types of it I talk to them as a group, about their general performance but I also talk to them individually. There are some tasks in which I gave them two or three minutes to each one in which I can talk to them about their progress every time they have reached an evaluation. I give them feedback. Sometimes it is written sometimes I talk to them about it. They reflect on what they do. They are aware of what they did and do so. And I think that it's good for them to know. Some teachers they have had. I'm not a keen at this point. I mean they have told me that. So I think that they really appreciate the way in which I am able to tell them what they really need to know about their progress.

**Interviewer:** How do you report your students' progress to your institution?

**LI-T4:** Well I have to give them the numbers. I mean what they got from the exam which represents the 50 percent and both together the speaking on the written exam. And you also have to let my coordinator know about the results from the portfolio the portfolio has to include five different tasks that are mandatory. I can ask my students to do more than five. But those are like extra practice. But they need to have five tasks and my coordinator wants to know or needs to know about the names of the tasks they contained, their aim of each one of them and that results

**Interviewer:** Is there a language skill you assess more than the others?

**LI-T4:** Yes. Well actually I assess three of them a lot, it's grammar. One of them which is maybe not one of the four skills but it's including all of them learn from them. Yes. And he also says speaking in writing which are the productive ones a lot.

**Interviewer:** All right. How do you make sure you follow specific criteria to assess your students with regard to productive skills?

**LI-T4:** Well I need to know what the requirements are. I take into consideration what my material is showing. I mean the materials we use includes a list of all the aspects students needs to develop and it's to prove in order to show that they're learning. And I also consulted the European framework and I also checked some of the articles that I can read sometimes about the topics in which I can make sure about it.

**Interviewer:** What can be some limitations or problems for your assessments in the language classroom?

**LI-T4:** Well I think the first problem when you have students that are supposed to be the same level but they are not performing the same way. So sometimes feedback even when it is supposed to help them it is discouraging them. Some people understand a feedback or assessment. If you are criticizing or judging them. So I need to be very careful in the way I give it to my students so they know that I am really worried about their progress that I am really interested in their learning process and that I am not trying to discourage them but the opposite I'm trying to encourage them to finish their course in the best way and to get the best results they can.

**Interviewer:** What do you think your students' needs in terms of language testing and glass based language assessment in order to improve the quality of learning?.

**LI-T4:** I think that the quality of learning in the institution is good and I think that the students take advantage of it. They really take advantage of what the institution offers and what the class

is about and what the materials include. This thing here is that sometimes they are under pressure and they pay more attention to aspects that are not providing them any help. I mean sometimes they are distracted with other things or they get lost in the process. So what I believe is that they need to focus more on what they have to learn. And I would like to have something like standardize a list of evidence. Is there something like that. I mean I really appreciate the freedom I have to decide which tasks I'm going to make my students perform. But I will also like to have a reference from my institution in which I can see and they can see in a clearer way what they are supposed to do.

**Interviewer:** So now what do you think you personally you need in terms of language testing in the classroom based language suddenly improve your teaching quality?

**LI-T4:** I consider that I have to use rubrics more than I do. Sometimes I don't give them enough time to be answered even when they are supposed to be easier because they come in in a list but sometimes I just can't take it or I ask them to answer me. But I am not into the process a lot. So I think that rubrics need to be something like a priority in my classes.

**Interviewer:** All right. Is there anything else you want to share with me about what you do in the classroom for a assessment practices

**LI-T4:** my years of experience as a teacher but I also had experience before as a student, and when I was a student I didn't receive good feedback from my teachers or my classmates. And I remember that sometimes I did it wrong and I didn't know why or how and I didn't know what to do. So I tried to help my students in that point in which I am really helping them know about their protests about their problems or issues they may have and how to correct them. So I am trying to change things in a way I don't know if I am really doing it but I think I am in the process of.

**Interviewer:** All right. Thank you very much for your time. So the end of the interview

## Appendix G Syllabus of the Workshop

The purpose of this course is to increase the participants' profile in the area of language testing and assessment. This course can be taken by pre-service and in-service language teachers who desire to expand their assessment knowledge, technical skills, classroom practices and their experiences to become language assessment literates. Participants **MUST** have access to students since they need to create, implement and analyse a language test.

General Aims:

By the end of this course, participants will be able to...

- Identify the usefulness of tests for classroom purposes
- Apply different methods to assess language in the classroom
- Analyse and critically evaluate language tests and items
- Recognise the importance of tests (learning evidences) to make judgements on the teaching practice.
- Become critical on the importance of educational measurement in language learning
- Identify problems in assessment and bring solutions to maximise their teaching practice.

This is a 20 hour course for classroom language teachers of English and it has been divided into 10 sessions. Each session last 2 hours, giving a total of 20 hours of instruction. Participants will also have to complete several practical and reflective tasks that include 6 hours of autonomous learning.

### OVERVIEW OF THE WORKSHOP

Session	Sessions	Contents	Activities
1 (Nov. 12)	Introduction of the course	Intervention outline, purpose of the course and important information	Pre-measure (Quiz)
2 (Nov. 13)	Testing, assessment and evaluation in language education	Definition of important concepts: testing, assessment, measurement, assessment literacy and evaluation Introducing and defining testing principles	Reflective Task 1
3 (Nov. 14)	Language testing	Development of language tests, types of responses (items), type of tasks and purpose of tests.	
4 (Nov 15)	Language Assessment Standards	Common European Framework for Language Reference, Formative and Summative assessments Standardised language Tests	
5 (Nov. 16)	Language assessment in receptive skills	Approaches to assess reading and listening Assessment for grammar and vocabulary in items	Practical Task 1
6 (Nov. 19)	Evaluation of language tests	Item analysis: item facility, item discrimination and distractor efficiency	
7 (Nov. 20)	Language assessment in productive skills	Approaches to assess writing and speaking skills use of holistic and analytic marking schemes	Reflective Task 2
8 (Nov. 21)	Communication of results and error correction	Use of feedback, error correction strategies, process of communication of assessment and results	
9 (Nov. 22)	Alternative Assessments		Practical Task 2

10 (Nov.23)	Language testing in Tamaulipas	Problems and solutions for the use of assessment in the language classroom	Post-measure (Quiz) Reflective Task 3
-------------	--------------------------------	--	--

### Session 1. Introduction to the Course

In this session, participants will become aware of the purpose and contents of the course. Instructor will provide the syllabus and answer questions from the participant related to how the course will be assessed and what participants can expect at the end. The administration of the Quiz A (pre-measure) will occur, this quiz is to measure how much you know at the beginning of the course. Instructor provides a list of materials and presents the reflective and practical tasks participants need to complete by the end of the course.

### Session 2. Testing, Assessment and Evaluation in Language Education

In this session, participants will discuss and define the concepts of *a test*, *measurement*, *assessment*, *evaluation* and *language assessment literacy*. Participants will have a presentation in which the instructor defines and asks participants about their previous knowledge of these concepts and the importance to distinguish their differences. In addition, this session introduces the foundations for the development of language tests. The testing principles (*validity*, *reliability*, *authenticity*, *practicality* and *washback*) are presented and defined with a theoretical and practical rationale. Participants will discuss the different types of validity and reliability, as well as, to discover to discover the authenticity of test and the practical aspects of these.

The participants have to complete the Reflective task 1 by the end of this session, this one can be written during the session or before session 3.

### Session 3. Language Testing

In this session, participants will identify the testing development cycle: from the identification of the purpose of the test to its final administration. An overview is provided in each aspect so, participants will be able to label certain characteristics related to the process of how to develop language tests or other testing instruments. Moreover, participants will be able to distinguish from a list of types of items their purpose and where they are appropriate to be use. In addition, different types of tests and their purpose for educational purposes will be discussed.

#### **Session 4. Language Assessment Standards**

In this session, participants will revise the levels of the CEFLR in order to use these as a level for reference to develop language tests. Besides this, the new CEFRL volume will be discussed to update teachers with this recently updated level. The types of assessment such as formative and summative will be discussed and also how teachers can balance these two types of assessment to empower the classroom practices. A revision of standardised test will be discussed to see the difference of the structure and language.

#### **Session 5. Language Assessment for Receptive Skills**

In this session, participants will discuss different approaches to assess the receptive skills (listening and reading). Participants will also discuss the assessment of grammar and vocabulary and how these two can be linked with the receptive skills. Participants need to pay attention to this because they need to find a strategy that leads to develop a section of a test. Participants need to complete the Practical task 1, they need to develop a task specification for the development of a test and design the test to be piloted and evaluated with items analysis (session 6).

#### **Session 6. Evaluation of Language Tests**

In this session, participants will explore the use of quantitative methods to analyse language tests and items. The methods from item facility, item discrimination and distractor efficiency will be discussed. As part of Practical task 1, participants will input the data with the methods they learnt to identify the efficiency of the items they developed.

#### **Session 7. Language assessment for Productive Skills**

In this session, participants will discuss different approaches to assess the productive skills (writing and speaking). Participants will also discuss the assessment of fluency and pronunciation and how these two can be linked with the productive skills. Participants will also discuss the different purposes of analytic and holistic marking schemes for these two language skills. Participants will also complete the Reflective Task 2.

#### **Session 8. Communication of results and error correction**

In this session, participants will discuss different methods for communicating classroom tests scores. The communication of assessment is divided into three phases, the informative (feedback), instruction and impact. Participants will review the efficiency of feedback in the

language classroom and critically evaluate other teachers' comments. In addition, participants will discuss different methods for error correction.

### **Session 9. Alternative Assessments**

In this session, participants will discuss different methods for assessing language apart from language tests. The use of different methods such as journals, essays, portfolios, rubrics, worksheets, peer-assessment, self-assessment among others, etc., will be discussed and participants need to complete Practical Task 2.

### **Session 10. Language Testing in Tamaulipas Classrooms**

In this session, participants will discuss about the use of language tests and other methods for language assessment that happen on participants' classrooms. Participants need to complete Reflective task 3 by the end of the session. Participants conclude this session with Quiz B and a refreshment session!

## **PRACTICAL AND REFLECTIVE TASKS**

### **Practical Tasks**

**Task 1:** Develop a task specification of one section of the test related to reading, listening or grammar and vocabulary,

- Develop adequate items for the section
- Administer the test with a small sample of test takers
- Input the data in Excel and conduct item facility, item discrimination and distractor analysis
- Interpret the results

**Task 2:** Evaluate a tool or an aspect which participants consider essential for assessment in the classroom

### **Reflective Tasks:**

1. Important considerations in language assessment for the classroom

- What is assessment for you and its importance in the language classroom?
- How do you conduct assessments in your classroom?
- What do you think the best methods for assessments are?

2. Development of tests and efficiency

- What elements do you consider essential for the development of language tests or testing instruments?

## Syllabus of the Workshop

- What are your perceptions towards the use of quantitative research methods to maximise language tests?
- What sort of information can you collect if you conduct item analysis?

### 3. Problems and solutions for Language assessment in the Tamaulipas classrooms of English

- What do you consider problematic at the moment of use assessments in your classroom?
- Can you provide one or more solutions to the problem(s) you have identified?

#### **Methods of Assessment**

There will not be an official assessment but you will be receiving constant feedback and the results of the quiz will be given at the end of the sessions.

#### **Course References List:**

Bachman, L. F. (2004). *Statistical Analysis for Language Assessment*. Cambridge: Cambridge University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Doff, A. (2004). *Teach English: A training Course for Teachers* (17<sup>th</sup> edition). Cambridge, Cambridge University Press.

Gower, R., Phillips, D., & Walters, S. (2005). *Teaching Practice: A Handbook for Teachers in Training*. Oxford: MacMillan.

Harmer, J. (2011). *How to Teach English* (17<sup>th</sup> edition). UK: Pearson Longman.

Hughes, A. (2013). *Testing for Language Teachers* (14<sup>th</sup> edition). Cambridge: Cambridge University Press.

McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.

Rea-Dickins, P., & Germaine, K. (1996). *Evaluation* (3<sup>rd</sup> edition). Oxford: Oxford University Press.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: principles and classroom Practice*. United Kingdom: Pearson Longman

Scrivener, J. (2011). *Learning Teaching: The Essential Guide to English Language Teaching* (3<sup>rd</sup> edition). Oxford: MacMillan.

## Appendix H Pre- and Post- Measure

### Pre and Post Measure

#### A. Match the statements with one of the main concepts of educational assessment based on its definition or purpose

- a) Assessment
- b) Test(ing)
- c) Measurement
- d) Teaching
- e) Evaluation

1. This term refers to the process of making decisions on students' performances.
2. Teachers use these tools to identify how much learners have improved during a particular course. These tools are used for several purposes to place students in different language proficiency levels or to identify success of achievement in students.
3. It is the process of developing tools, piloting tasks, marking and communicating results about students' performances.
4. It is the process of providing a number or quantity to a task or a testing tool based on how test-takers' performed.

#### B. Match the statements with one of the testing principles based on its definition or purpose

- Testing principles*
- f) Validity
  - g) Practicality
  - h) Washback (impact)
  - i) Authenticity
  - j) Reliability
  - k) Stability

5. It focuses on the use of resources, available budget, time and location for test administration.
6. The items from a test are consistent and can be administer to another sample of test takers.
7. The results can bring benefits or challenges for the classroom or outside the classroom.
8. A tests needs to measure what it is proposed to test.
9. Natural language, contextualised items, meaningful and organised topics to be included in a language test.

**C. Match the statements with one alternative assessments based on its definition or purpose**

- a) Essays and journals
- b) learning portfolio
- c) Presentations
- d) worksheets and workbook
- e) rubrics and checklist
- f) group and individual tasks
- g) self- and peer- assessments
- h) holistic rubric
- i) analytical rubric

- 10. It collects evidence of how the learning process has happened
- 11. This type of assessments are mainly for oral activities in which students should explain and prepare some visuals
- 12. Students evaluate their progress with these forms
- 13. Students have to develop a project for the class
- 14. a method of assessment to test students' criticality in an area
- 15. type of assessment that provides a mark in a range of linguistic features.

**D. Match the statements with one of the test types and purposes based on its definition or purpose**

- a) Achievement
- b) Diagnostic
- c) Placement
- d) Proficiency
- e) Attitude
- f) Ipsative
- g) progress

- 16. This type of test is used to assess owns personal progress
- 17. Test administered at the beginning of the course in the classroom
- 18. Institutional assessment to locate students' language level
- 19. Test mostly administer at the end of the course
- 20. Test to identify whether students have learnt what they were taught in a short time

**E. Match the statements with one of the types for item evaluation based on its definition or purpose**

- |                          |
|--------------------------|
| a) Item facility index   |
| b) Classical Test Theory |
| c) Distractor analysis   |
| d) Discrimination Index  |
| e) Item Response Theory  |
| f) Analysis of Raters    |

21. This is to revise the quality of the test among higher scores and lower scores
22. This theory relies on how items are analysed based on the level of difficulty and structure of the items
23. This analysis is useful to identify the quality of multiple choice items
24. Measures how difficult or easy the items from a test are.
25. Two marks of a speaking tests have significant discrepancies

**F. Match the statements with one of the types for communication of assessments and results based on its definition or purpose**

- |                  |
|------------------|
| a) Informative   |
| b) Instructional |
| c) Impact        |

26. Teacher informs students about tests
27. Teacher provides feedback to students
28. Teachers share with other colleagues their students' scores
29. Teacher highlights the important contents to revise
30. Language programme managers uses the students' scores to make decisions

**G. Match the statements with one of the types of responses based on its definition or purpose**

- |   |
|---|
| a) Selected responses<br>b) Productive responses<br>c) Personal responses |
|---|

- 31. Cloze items
- 32. Matching items
- 33. True, false, not given items
- 34. Multiple –Choice items
- 35. Fill-in-the gap items

**H. Match the statements with one of the testing processes based on their definitions or purposes.**

- |   |
|---|
| a) Pilot<br>b) Tasks specifications<br>c) Evaluation of items<br>d) Purpose and type of a test<br>e) Assessment as an educational support<br>f) Performance assessments |
|---|

- 36. Teachers develop a document where they describe each stage of a test; including the content, abilities, types of items, time and level
- 37. This stage is important at the moment of developing classroom tests to increase the validity
- 38. Tests can guide teachers how well they teach and develop their testing sources
- 39. Types of assessments that do not include tests.
- 40. Teachers and language programme managers can revise test scores to highlight assessment needs

## Responses

1	E	21	D
2	B	22	B
3	A	23	D
4	C	24	A
5	G	25	F
6	J	26	A
7	H	27	B
8	F	28	C
9	I	29	A
10	B	30	C
11	C	31	B
12	G	32	A
13	E	33	A
14	A	34	A
15	I	35	B
16	F	36	B
17	B	37	A
18	C	38	E
19	A	39	F
20	G	40	E

## Appendix I Example of Practical and Reflective Tasks

For me assessment is the way to see if my students learnt or not what I taught during the class. It is very important because it helps us identify students who need more help or more preparation. I usually have tests, worksheets, pages of the workbook, attendance and class participation. The test is 30% of each partial and 70% the other aspects. I think Ss like communicative tasks, but it is very difficult to assess sometimes. I have 40 students, and sometimes is very crazy to see each student. I try to give worksheets to see the progress of my students.

## Appendix J Workshop Evaluation Sheet

### Instructions:

This Course Evaluation Sheet is divided into two parts. *Part A* contains a total of 20 items categorized into 5 sections: items from 1 to 4 refers to the objectives and contents of the course, 5 to 8 from materials and resources applied and used, 9 to 12 from participation and tasks of the course, 13 to 17 instructor's performance and 18 to 20 from practicality of the training. Items from part A (1 to 20) need to be answered based on a Likert-scale from 1 to 5 (1= strongly disagree, 2= disagree, 3=neither disagree nor agree, 4= agree and 5=strongly agree). *Part B* includes 5 open-ended questions related to the overall course evaluation.

### PART A

Statements	1	2	3	4	5
1. The objectives of the training were clearly cover					
2. The objectives of the tasks were related to the contents of the course					
3. The content of the course was relevant to my teaching practice					
4. The content was structured and delivered adequately					
5. The materials distributed were helpful					
6. The materials were connected to the discussions and presentation of content					
7. The materials covered too much they were not clear enough					
8. The contextualized course met the current needs in my institution					
9. Participation and discussion were encouraged					
10. The tasks were explained in detail to obtain better outcomes.					
11. The list of references was useful to understand the structure of the course.					
12. Presentations, videos, tasks and journals were beneficial for my learning.					
13. The instructor used a different materials and resources to explain the subject.					
14. The instructor stimulated my interests to complete the course					
15. The instructor developed a large number of activities					
16. The instructor provided materials to read and be prepared for each session.					
17. The instructor motivated me throughout a diversity of tasks and activities					
18. The length of time for instruction and tasks was enough					
19. The course was useful for my professional development					
20. The tasks were suitable for the purpose of professional development					

**PART B**

21. What was the most useful?

.....  
.....  
.....

22. What was the least useful?

.....  
.....  
.....

23. What else would you like to see included in this training? Are there any other topics that you would like to be offered?

.....  
.....  
.....

24. Would you recommend this course to colleagues? Yes/No Why?

.....  
.....  
.....

25. Any other comments?

.....  
.....  
.....

## **Appendix K      Focus Group Interview Guidelines**

1. What is your overall perspective of the course?
2. To which extent the materials from the course were useful?
3. After taking the course, what are the aspects you have changed to improve assessment in the classroom?
4. What practical and/or reflective tasks were/was more significant to you?
5. After the course, has assessment become easier to be conducted in the classroom? Why?
6. After the course, do you consider that your assessment skilfulness and practices have improved after the course?
7. Which areas of the course do you think you improved the most?
8. Which areas of the course do you think you improved the less?
9. What changes have you done to improve your assessment skilfulness and practices in the classroom?
10. If you have the opportunity to change something from the course, what would it be and why?
11. Would you recommend this course to other colleagues? Why?
12. Is there anything you would like to mention about the impact of the course into your teaching practice? Elaborate...

## Appendix L Findings from Participants' Educational Sectors

Components of the Measure	Public Sector (n=4)									
	Pre- Measure					Post- Measure				
	Min	Max	SUM	Mean	SD	Min	Max	SUM	Mean	SD
Concepts in assessment	2	3	10	2.5	0.58	3	3	12	3	0
Testing principles	3	4	14	3.5	0.58	4	5	17	4.25	0.5
Types of items and tests	3	5	17	4.25	0.96	6	8	28	7	0.82
Development of Testing instruments	2	3	11	2.75	0.5	5	6	21	5.25	0.5
Marking Schemes	3	3	12	3	0	3	3	12	3	0
Item analysis	0	3	3	0.75	0.5	3	4	15	3.75	0.5
Communication of Results	1	2	7	1.75	0.5	3	3	12	3	0
Alternative assessments	2	4	13	3.25	0.96	5	5	20	5	0
Educ. Assess	2	3	11	2.75	0.5	3	3	12	3	0
<b><u>Overall</u></b>			<b><u>98</u></b>	<b><u>24.5</u></b>	<b><u>3.51</u></b>			<b><u>149</u></b>	<b><u>37.25</u></b>	<b><u>0.96</u></b>

Components of the Measure	Private Sector (n=3)									
	Pre- Measure					Post- Measure				
	Min	Max	SUM	Mean	SD	Min	Max	SUM	Mean	SD
Concepts in assessment	2	3	7	2.33	0.58	2	3	9	3	0
Testing principles	3	4	11	3.67	0.58	5	5	15	5	0
Types of items and tests	3	5	12	4	1	7	8	22	7.33	0.58
Development of Testing instruments	4	6	14	4.67	1.15	5	6	17	5.67	0.58
Marking Schemes	3	3	9	3	0	3	3	9	3	0
Item analysis	0	1	2	0.67	0.58	4	4	12	4	0
Communication of Results	2	3	7	2.33	0.58	3	3	9	3	0
Alternative assessments	3	4	11	3.67	0.58	5	5	15	5	0
Educ. Assess	1	2	4	1.33	0.58	3	3	9	3	0
<b><u>Overall</u></b>			<b><u>78</u></b>	<b><u>26</u></b>	<b><u>3.46</u></b>			<b><u>117</u></b>	<b><u>39</u></b>	<b><u>0</u></b>

Components of the Measure	Language Institute Sector(n=3)									
	Pre- Measure					Post- Measure				
	Min	Max	SUM	Mean	SD	Min	Max	SUM	Mean	SD
Concepts in assessment	2	3	7	2.33	0.58	3	3	9	3	0
Testing principles	2	4	9	3	1	4	5	14	4.67	0.58
Types of items and tests	4	6	15	5	1	7	8	22	7.33	0.58
Development of Testing instruments	4	5	14	4.67	0.58	5	6	17	5.67	0.58
Marking Schemes	3	3	9	3	0	3	3	9	3	0
Item analysis	1	2	4	1.33	0.58	4	4	12	4	0
Communication of Results	2	3	7	2.33	0.58	3	3	9	3	0
Alternative assessments	5	5	15	5	0	5	5	15	5	0
Educ. Assess	2	3	8	2.67	0.58	3	3	9	3	0
<b><u>Overall</u></b>			<b><u>88</u></b>	<b><u>29.33</u></b>	<b><u>0.58</u></b>			<b><u>116</u></b>	<b><u>38.67</u></b>	<b><u>1.15</u></b>

## Appendix M Participants' Individual Scores from Pre- and Post- Measure

Sector	Concepts assessment (3)			Testing principles (5)			Types of items & tests (8)			Dev. Of Testing instruments (6)			Marking Schemes (3)			Comm. of Results (3)			Alternative assessments (5)			Item analysis (4)			Educ. Assess (3)			Totals			
	B	E	D	B	E	D	B	E	D	B	E	D	B	E	D	B	E	D	B	E	D	B	E	D	B	E	D	B	E	D	
PUB1	2	3	1	3	4	1	4	7	3	3	5	2	3	3	0	1	3	2	2	5	3	0	4	4	3	3	0	21	37	16	
PUB2	3	3	0	4	4	0	5	8	3	3	6	3	3	3	0	2	3	1	4	5	1	0	3	3	3	3	3	0	27	38	11
PUB3	2	3	1	3	5	2	3	7	4	2	5	3	3	3	0	2	3	1	3	5	2	2	4	2	2	3	1	22	38	16	
PUB4	3	3	0	4	4	0	5	6	1	3	5	2	3	3	0	2	3	1	4	5	1	1	4	3	3	3	0	28	36	8	
<b>(SUM)</b>	<b><u>10</u></b>	<b><u>12</u></b>	<b><u>2</u></b>	<b><u>14</u></b>	<b><u>17</u></b>	<b><u>3</u></b>	<b><u>17</u></b>	<b><u>28</u></b>	<b><u>11</u></b>	<b><u>11</u></b>	<b><u>21</u></b>	<b><u>10</u></b>	<b><u>12</u></b>	<b><u>12</u></b>	<b><u>0</u></b>	<b><u>7</u></b>	<b><u>12</u></b>	<b><u>5</u></b>	<b><u>13</u></b>	<b><u>20</u></b>	<b><u>7</u></b>	<b><u>3</u></b>	<b><u>15</u></b>	<b><u>12</u></b>	<b><u>11</u></b>	<b><u>12</u></b>	<b><u>1</u></b>	<b><u>98</u></b>	<b><u>149</u></b>	51	
PRIV1	2	3	1	4	5	1	3	7	4	4	6	2	3	3	0	2	3	1	4	5	1	0	4	4	2	3	1	24	39	15	
PRIV2	3	3	0	3	5	2	4	7	3	4	6	2	3	3	0	2	3	1	3	5	2	1	4	3	1	3	2	24	39	15	
PRIV3	2	3	1	4	5	1	5	8	3	6	5	-1	3	3	0	3	3	0	4	5	1	1	4	3	2	3	1	30	39	9	
<b>(SUM)</b>	<b><u>7</u></b>	<b><u>9</u></b>	<b><u>2</u></b>	<b><u>11</u></b>	<b><u>15</u></b>	<b><u>4</u></b>	<b><u>12</u></b>	<b><u>22</u></b>	<b><u>10</u></b>	<b><u>14</u></b>	<b><u>17</u></b>	<b><u>3</u></b>	<b><u>9</u></b>	<b><u>9</u></b>	<b><u>0</u></b>	<b><u>7</u></b>	<b><u>9</u></b>	<b><u>2</u></b>	<b><u>11</u></b>	<b><u>15</u></b>	<b><u>4</u></b>	<b><u>2</u></b>	<b><u>12</u></b>	<b><u>10</u></b>	<b><u>5</u></b>	<b><u>9</u></b>	<b><u>4</u></b>	<b><u>78</u></b>	<b><u>117</u></b>	39	
LI1	2	3	1	3	5	2	6	7	1	5	5	0	3	3	0	2	3	1	5	5	0	1	4	3	3	3	0	30	38	8	
LI2	3	3	0	2	4	2	5	7	2	5	6	1	3	3	0	3	3	0	5	5	0	1	4	3	2	3	1	29	38	9	
LI3	2	3	1	4	5	1	4	8	4	4	6	2	3	3	0	2	3	1	5	5	0	2	4	2	3	3	0	29	40	11	
<b>(SUM)</b>	<b><u>7</u></b>	<b><u>9</u></b>	<b><u>2</u></b>	<b><u>9</u></b>	<b><u>14</u></b>	<b><u>5</u></b>	<b><u>15</u></b>	<b><u>22</u></b>	<b><u>7</u></b>	<b><u>14</u></b>	<b><u>17</u></b>	<b><u>3</u></b>	<b><u>9</u></b>	<b><u>9</u></b>	<b><u>0</u></b>	<b><u>7</u></b>	<b><u>9</u></b>	<b><u>2</u></b>	<b><u>15</u></b>	<b><u>15</u></b>	<b><u>0</u></b>	<b><u>4</u></b>	<b><u>12</u></b>	<b><u>8</u></b>	<b><u>8</u></b>	<b><u>9</u></b>	<b><u>1</u></b>	<b><u>88</u></b>	<b><u>116</u></b>	28	

Note: B represents (pre- measure), E represents (post- measure) and D represents (difference); the number in parenthesis at the top of each column represents the number of items in each assessment component of the measures.



## Appendix N Language Assessment Literacy Development Framework (Descriptors)

Theoretical Framework Components	Stages for Language Assessment Literacy Development				
	Illiteracy	Nominal	Functional	Procedure and Conceptual	Multidimensional
<b>Applied Linguistics</b>	Has minimal knowledge of the use and instruction of language in the classroom. Minimal knowledge of language education theories and the application of these in the language teaching practice	Has basic knowledge of the use and instruction of language in the classroom. Basic knowledge of language education theories and the application of these in the language teaching practice	Has a reasonable understanding of the use and instruction of language in the classroom. Reasonable knowledge of language education theories and the application of these in the language teaching practice	Has a relevant understanding and knows how to use and instruction of language in the classroom. Relevant knowledge of language education theories and the application of these in the language teaching practice	Has a sophisticated understanding and use of the use and instruction of language in the classroom. Sophisticated knowledge of language education theories and the application of these in the language teaching practice
<b>Educational Assessment</b>	Has minimal knowledge of assessment of/for/ as learning. The practices of assessment in the classes are minimal and there is no evidence of tracking students' achievement	Has basic knowledge of assessment of/for/ as learning. The practices of assessment in the classes are basic and there is little evidence of tracking students' achievement	Has a reasonable understanding of assessment of/for/ as learning. The practices of assessment in the classes are appropriate and there is evidence of tracking students' achievement	Has a relevant understanding and knows how to assessment of/for/ as learning. The practices of assessment in the classes are relevant and there is appropriate evidence of tracking students' achievement	Has a sophisticated understanding and use of assessment of/for/ as learning. The practices of assessment in the classes are sophisticated and there is a lot of evidence of tracking students' achievement
<b>Institutional assessment policies</b>	Has minimal knowledge of the local, regional and	Has basic knowledge of the local, regional and national assessment policies.	Has a reasonable understanding of the local,	Has a relevant understanding and knows how to apply the local, regional and national	Has a sophisticated understanding and historical review of the local, regional

Language Assessment Literacy Development Framework (Descriptors)

	national assessment policies.		regional and national assessment policies.	assessment policies among different teaching contexts.	and national assessment policies.
<b>Language Testing principles</b>	Has minimal knowledge of the testing principles and struggles to differ concepts related to testing, assessment and evaluation.	Has basic knowledge of the testing principles and can identify different components of testing, assessment and evaluation.	Has a reasonable understanding of the testing principles and can identify useful qualities between testing procedures, assessment stages and types of evaluation	Has a relevant understanding and knows how to use the testing principles, its qualities and applications towards improving language tests. Can use different testing procedures, assessment stages and types of evaluation	Has a sophisticated understanding and use of the testing principles. Can apply different methods to identify whether tests can be reliable, valid, as well as measuring the impact, practicality and authenticity of language tests. Also, has enough understanding of the advantages of the testing principles to identify different stages for assessment practices and evaluation.
<b>Language Testing and Assessment</b>	<p>Has minimal knowledge of the use and application of the CEFR and lack of reference to other standardised criteria for language education.</p> <p>Has minimal experience on using different types of tests, items and tasks for specific language skill.</p>	<p>Has basic knowledge of the use and application of the CEFR and know a few reference to other standardised criteria for language education.</p> <p>Has basic experience on using different types of tests, items and tasks for specific language skill.</p>	<p>Has a reasonable knowledge of the use and application of the CEFR and other standardised criteria from different frameworks for language education.</p> <p>Has reasonable experience on using different types of tests, items and tasks for specific language skill</p>	<p>Has a relevant knowledge of the use and application of the CEFR and other well-known local and international standardised criteria from different frameworks for language education.</p> <p>Has relevant experience on using different types of tests, items and tasks for specific language skill</p>	<p>Has a sophisticated knowledge of the use and application of the CEFR and other well-known local and international standardised criteria from different frameworks for language education.</p> <p>Can make comparisons using different scales to measure proficiency</p> <p>Has sophisticated experience on using different types of tests for classroom purpose, items and tasks for specific language skill or linguistic element</p>

<p><b>Quality Assurance in language tests</b></p>	<p>Has minimal knowledge and experience on conducting item analysis on the classroom testing instruments.</p>	<p>Has basic knowledge and experience on conducting item analysis on the classroom testing instruments.</p>	<p>Has reasonable knowledge and experience on conducting item analysis on the classroom testing instruments.</p> <p>Uses the test scores to identify different higher and lower test-takers' scores and can identify several inaccuracies on teaching, learning and assessing practices</p>	<p>Has relevant knowledge and experience on conducting item analysis on the classroom testing instruments.</p> <p>Uses the test scores to identify and compare different higher and lower test-takers' scores and can identify and prevent inaccuracies on teaching, learning and assessing practices.</p> <p>Can modify testing instruments based on the psychometric properties of the tests.</p>	<p>Has sophisticated knowledge and experience on conducting item analysis on the classroom testing instruments.</p> <p>Uses the test scores to identify compare and evaluate different higher and lower test-takers' scores and can easily identify, prevent and modify inaccuracies on teaching, learning and assessing practices.</p> <p>Can apply different approaches to modify testing instruments based on the psychometric properties of the tests.</p>
<p><b>Assessment purposes</b></p>	<p>Has minimal knowledge on the different aspects that assessment can provide to students and teachers.</p>	<p>Has basic knowledge on the different aspects that assessment can provide to students and teachers.</p> <p>Can identify testing instruments for diagnostic, formative and summative.</p>	<p>Has reasonable knowledge on the different aspects that assessment can provide to students and teachers.</p> <p>Can identify and modify testing instruments for diagnostic, formative and summative.</p>	<p>Has relevant knowledge on the different aspects that assessment can provide to students and teachers.</p> <p>Can adapt and apply testing instruments for diagnostic, formative and summative.</p> <p>Provides lessons in which the assessment practices are relatively balanced</p>	<p>Has sophisticated knowledge on the different aspects that assessment can provide to students and teachers.</p> <p>Can develop and create testing instruments for specific purposes like diagnostic, formative and/or summative.</p> <p>Provides lessons in which the assessment practices are completely balanced</p>

Language Assessment Literacy Development Framework (Descriptors)

<p><b>Alternative assessments</b></p>	<p>Has minimal knowledge and experience of using different assessment tools for the language classroom.</p> <p>Relies on standardised tests to measure students' progress.</p> <p>Has difficulties to find out sources that benefit students' progress.</p>	<p>Has basic knowledge and experience of using different assessment tools for the language classroom.</p> <p>Relies on standardised tests to measure students' progress but sometimes uses other sources like classroom tests.</p> <p>Has minimal difficulties to find out sources that benefit students' progress.</p>	<p>Has reasonable knowledge and experience of using and applying different assessment tools for the language classroom.</p> <p>Relies on different learning evidences to measure students' progress.</p> <p>Has no difficulties to find out sources that benefit students' progress.</p>	<p>Has relevant knowledge and experience of using, selecting and applying different assessment tools for the language classroom.</p> <p>Relies on a variety of learning evidences to measure students' progress.</p> <p>Has no difficulties to find out and adapt new sources that benefit students' progress in relation to their learning process.</p>	<p>Has sophisticated knowledge and experience of using, selecting and applying different assessment tools for the language classroom.</p> <p>Relies on different elements that allows to have a critical perspective on students' process with the learning evidences to measure students' progress.</p> <p>Has no difficulties to find out and adapt new sources that benefit students' progress in relation to their learning process.</p>
<p><b>Communication of Assessments</b></p>	<p>Has minimal knowledge and experience on how to communicate test results to different stakeholders.</p> <p>Provides with minimal information to students about the contents of the tests and how this will be assessed students' knowledge.</p>	<p>Has basic knowledge and experience on how to communicate test results to different stakeholders.</p> <p>Provides basic information to students about the test contents and how they will be assessed.</p> <p>Provides basic feedback to students' performances.</p>	<p>Has reasonable knowledge and experience on how to communicate test results to different stakeholders.</p> <p>Provides basic information to students about the test contents and how they will be assessed before every testing situation.</p> <p>Provides relevant holistic and analytic feedback to</p>	<p>Has relevant knowledge and experience on how to communicate test results to different stakeholders.</p> <p>Provides enough information to students about the test contents and how they will be assessed before every testing situation.</p> <p>Provides enough holistic and analytic feedback to students' performances; either verbally or</p>	<p>Has sophisticated knowledge and experience on how to communicate test results to different stakeholders.</p> <p>Provides sufficient information to students about the test contents and how they will be assessed before every testing situation.</p> <p>Provides sufficient holistic and analytic feedback to students' performances; either verbally</p>

	<p>Provides minimal feedback to students' performances.</p> <p>Lacks of sharing information about students' scores with other teachers and institution authorities</p>	<p>Shares little information about students' scores to other teachers and institution authorities</p>	<p>students' performances; either verbally or written.</p> <p>Shares reasonable information about students' tests' scores to other teachers and institution authorities to compare and be aware of differences on methods for teaching, learning and assessing language.</p>	<p>written and keeps a record of individual students' progress</p> <p>Shares enough information about students' tests' scores to other teachers and institution authorities to compare and be aware of different methods for teaching, learning and assessing language.</p>	<p>or written and keeps a record of individual students' progress.</p> <p>Shares sufficient information about students' tests' scores to other teachers and institution authorities to identify, compare and become critical on the different methods for teaching, learning and assessing language.</p>
<b>Assessment processes</b>	<p>Has minimal knowledge and experience of developing assessments for the language classroom.</p> <p>Finds with a high degree of difficulty to identify suitable instruments for the purpose of the test.</p>	<p>Has basic knowledge and experience of developing assessments for the language classroom.</p> <p>Finds with certain difficulty to identify suitable instruments for the purpose of the test.</p>	<p>Has reasonable knowledge and experience of developing assessments for the language classroom.</p> <p>Knows the stages of test development and to create test specifications</p> <p>Finds with minimal difficulty to identify suitable instruments for the purpose of the test.</p>	<p>Has relevant knowledge and experience of developing assessments for the language classroom.</p> <p>Knows each of the stages of test development and is able to create test specifications and items for the tasks</p> <p>Finds with no difficulty to identify suitable instruments for the purpose of the test.</p>	<p>Has sophisticated knowledge and experience of developing assessments for the language classroom.</p> <p>Clearly knows each stage and can provide useful information for the test development and is able to create test specifications and items for the tasks</p> <p>Can identify, with no hesitation, suitable instruments for the purpose of different tests.</p>



## List of References

- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do? In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching*, 205–221. <http://dx.doi.org/10.1007/s10972-012-9278>.
- Alberta Educational Institute (2019). <https://education.alberta.ca/literacy-and-numeracy/literacy/everyone/what-is-literacy/>
- Alderson, J. C., & Wall, D. (1993) Does washback exist? *Applied Linguistics* 14, 115-129.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- American Federation of Teachers, National Council on Measurement in Education, & National Educational Association (1990). Standards for Teacher Competence in Educational Assessment of Students. Washington DC: American Federation of Teachers. Available online: <http://www.unl.edu/buros/bimm/html/article3.html>.
- Araujo, N. (2006). *Teacher assessment practices at the centre for languages and applied linguistics of the University of Tamaulipas in Mexico* (Master Dissertation). University of Essex, United Kingdom.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Bachman, L., & Kunnan, A. (2005). *Statistical Analyses for Language Assessment: Workbook and CD ROM*. Cambridge: Cambridge University Press.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey, K. M., & Brown, J. D. (1996). Language testing courses: What are they? In Cumming, A., and Barting, I., Martin, M. and Vedder, I. (Eds.) (2010). Communicative proficiency and linguistic development: intersections between SLA and language testing research. Euroasla Monograph 1. Available online: <http://eurosla.org/monographs/EM01/EM01home.html>.

## List of References

- Barnes, N., Fives, H., & Dacey, C. M. (2017). US teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education, 65*, 107-116.
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science education, 85*(5), 536-553.
- Benigno, Veronica & De Jong, John HAL & Van Moere, Alistair. (2017). How long does it take to learn a language? Insights from research on language learning.
- Berry, V., & O'Sullivan, B. (2016, April). Assessment Literacy for language teachers. Paper presented at the TEA SIG PCE.
- Berry, V., Sheehan, S., & Munro, S. (2017) Exploring teachers' language assessment literacy: a social constructivist Approach to Understanding Effective Practices. In: *ALTE 6<sup>th</sup> International Conference- Learning and Assessment: Making connections*, 3-5 May 2017, Bologna, Italy.
- Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan, 80*.
- Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language teaching research, 10*(1), 3-31.
- Borich, G.D., & Tombari, M.L. (2004). *Educational assessment for the elementary and middle school classroom*. New Jersey: Pearson Education Inc.
- Boyles, P. (2005). Assessment literacy. In M. Rosenbusch (Ed.), *National Assessment Summit Papers*. (pp. 1-15). Ames, IA: Iowa State University.
- Brindley, G. (2001). Language assessment and professional development. In Elder, C., Brown, A., Grove, E., Hall, K., Iwashita, N., Lumley, T., McNamara, T., and O'Loughlin, K. (Eds.) *Experimenting with Uncertainty. Essays in Honour of Alan Davies*. Cambridge: Cambridge University Press, 126 – 136.
- Brindley, G. (2008). Educational reform and language testing. In Shohamy, E. and Hornberger, N. (Eds.) *Encyclopedia of Language and Education: Vol. 7, Language Testing and Assessment*. New York: Springer, 365 - 378.
- British Council. (2015). *Continuing Professional Development (CPD) Framework for Teachers*: British Council.

- British Council. (May 2015). English in Mexico: An examination of policy, perceptions and influencing factors. British Council Report.
- Brookhart, S. M. (1998). Teaching about grading and communicating assessment results. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA. (ERIC Document Reproduction Service No. 419838)
- Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, 18(1), 5-13.
- Brookhart, S. M. (2001). The Standards and classroom assessment research. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas, TX. (ERIC Document Reproduction Service No. ED451189)
- Brown, H. D. (2006). *Language Assessment: Principles and Classroom Practice*. New York: Pearson.
- Brown, H. D. (2007). *Principles of language learning and teaching*. Longman Pearson: London.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. New York: Pearson Longman.
- Brown, J. D. (2001). *Using Surveys in Language Programs*. Cambridge: Cambridge University Press.
- Brown, J. D. (2005). *Testing in Language Testing Programs: A Comprehensive Guide to English Language Assessment*. New York: McGraw-Hill.
- Brown, J. D. (2010). How are PCA and EFA used in language test and questionnaire development? *Shiken* 14(2), 30 – 35.
- Brown, J. D. (2012). Choosing the right type of assessment. In C. Coombe, P. Davison, B. O'Sullivan, & S. Stoyhoff (eds.), *The Cambridge Guide to Second Language Assessment*, 133-139, Cambridge: Cambridge University Press.
- Brown, J. D., & Bailey, K. M. (2008). Language testing courses: what are they in 2007? *Language Testing* 25(3) 349 – 384.
- Brumfit, C. (2010). Literacy or literacies? Academic identities in a language –sharing world. In G. Blue (Ed.), *Developing academic literacy: (pp.13-24)* Bern. Switzerland: Peter Lang.

## List of References

- Brumfit, C. J., & Johnson, K. (1979). *The communicative approach to language teaching*. Oxford: Oxford University Press.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Cheng, L. (2008). Washback, Impact and Consequences. In Shohamy, E. and Hornberger, N. H. (Eds.) *Encyclopedia of Language and Education*. New York: Springer, 349 – 364.
- Cohen, A. D. (1994). *Assessing language ability in the classroom*. Boston: HH.
- Cohen, L., Manion, L. & Morrison, K., (2013) *Research methods in education*. Routledge.
- Coniam, D. (2009). Investigating the quality of teacher-produced tests for EFL students and the effects of training in test development principles and practices on improving test quality. *System*, 37(2), 226-242.
- 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.
- Creswell, J.W. (2003) *Research design: Qualitative, quantitative, and mixed methods approaches*, 2nd edn. Thousand Oaks, CA: SAGE.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1). Retrieved from <http://epaa.asu.edu/epaa/vol8.html>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing* 25(3) 327 – 348.
- Davies, P., & Pearse, E. (2000). *Success in English language teaching*. Oxford: Oxford University Press.
- Davison, C. & Leung, C (2009). Current issues in English language teacher-based assessment. *Journal TESOL* 43(3), 393-415.

- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334.
- Davison, C., & Hamp-Lyons, L. (2010). 'The Hong Kong Certificate of Education: school-based assessment reform in Hong Kong English language education', in L.Cheng and A. Curtis (eds.), *English Language Assessment and the Chinese Learner*. New York, NY: Routledge. 248–62.
- 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.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a) Approaches to classroom assessment inventory: A new Instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248-266, DOI:[10.1080/10627197.2016.1236677](https://doi.org/10.1080/10627197.2016.1236677).
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b) Teacher assessment literacy: A review of international standards and measures. *Educational Assessment*. 28:251-272
- Díaz de la Garza, A. M. E. (2017). The emerging identity of pre-service teachers during the practicum component of a second language teacher education programme at a Mexican State university, in Grounds, P. and Moore, C (eds.) *Higher Education English Language Teaching and Research in Mexico*. Mexico: British Council and University of Southampton: pp: 49-72.
- Dörnyei, Z. (2007) *Research methods in applied linguistics*: Oxford: Oxford University Press.
- Dunkley, D. (2004). Current practices in the testing of oral communication: A case study in a Japanese university. *Aichi Gakuin General Studies Journal*, 51, (3)73-80.
- Dunkley, D. (2014). Britain's new language testing powerhouse: An interview with Professor Tony Green. *The Language Teacher*, 45(4), 13-16.
- Eames, C. (2003). Researching in cooperative education: How a practitioner met the challenge. In *Handbook for research in cooperative education and internships* (pp. 83-106). Routledge.
- Earl, L (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin.
- Eckhout, T., Davis, S., Mickelson, K., & Goodburn, A. (2005). A method for providing assessment training to in-service and pre-service teachers. Paper presented at the annual meeting of the Southwestern Educational Research Association, New Orleans, LA.

## List of References

- Ellis, R. (1995). Appraising second language acquisition theory in relation to language pedagogy. In Cook, G., Seidlhofer, B. (eds.) *Principle and Practice in Applied Linguistics*. Oxford: Oxford University Press.
- Extra, G., Spotti, M., & Van Avermaet, P. (Eds.) (2009). *Language Testing, Migration and Citizenship: Cross-National Perspectives on Integration Regimes*. London: Continuum.
- Falsgraf, C. (2005). Why a national assessment summit? New visions in action. *National Assessment Summit*, p. 6.
- Fertig, M. (2003). *Who's to Blame? The Determinants of German Students' Achievement in the Field*. London: SAGE Publications.
- Field, A. (2000). *Discovering statistics using IBM SPSS Statistics*. London: SAGE Publications.
- Frawley, W. (1989). Thought and Language. Lev Vygotsky (A. Kozulin, rev. ed.). Cambridge, MA: MIT Press, 1986. Pp. lxi 287. *Studies in Second Language Acquisition*, 11(3), 331-332.  
doi:10.1017/S0272263100008172
- Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Education.
- Fulcher, G. (2009). Test use and political philosophy. *Annual Review of Applied Linguistics* 29, 3 - 20.
- Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132.
- Fulcher, G. (2013). Language Testing and Philosophy. In A. Kunnan (Ed.), *Companion to Language Assessment*. London: Wiley-Blackwell.
- Fulcher, G. and Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London and New York: Routledge.
- Gass, S. M., & Selinker, L. (2001). *Second language Acquisition: An introductory course*. Lawrence Erlbaum: London
- Gebese, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Giraldo, F. (2018b). Language assessment literacy: Implications for language teachers. *Profile Issues in Teachers Professional Development*, 20(1), 179-195.

- Giraldo, F., & Murcia, D. (2018a). Language Assessment Literacy for Pre-Service Teachers: Course Expectations from Different Stakeholders. *GiST Education and Learning Research Journal*, 16, 56-77.
- González, E.F. (2017). The challenge of EFL writing assessment in Mexican higher education, in Grounds, P. and Moore, C (eds.) *Higher Education English Language Teaching and Research in Mexico*. Mexico: British Council and University of Southampton: pp: 73-100.
- Gottheiner D. M., & Siegel M. A. (2012). Experienced Middle School Science Teachers' Assessment Literacy: Investigating Knowledge of Students' Conceptions in Genetics and Ways to Shape Instruction. The Association for Science Teacher Education, USA J Sci Teacher Educ 23, 531–557. <http://dx.doi.org/10.1007/s10972-012-9278z>
- Green, A. (2013). *Exploring language assessment and testing: Language in action*. Routledge.
- Grounds, P., & Moore, C. (2017). Higher education English language teaching and research in Mexico. British Council Mexico
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing* 14(3), 295-303.
- Han, Z., & Anderson, N. (Eds.) (2009). *Second language reading research and instruction*. Ann Arbor, MI: University of Michigan Press.
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development.
- Harrison, A. (1983). *A language testing handbook*. London: McMillan.
- Hasselgreen, A., Carlsen, C., & Helness, H. (2004). European Survey of Language Testing and Assessment Needs. Part 1: General Findings. EALTA. Available online: <http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf>.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Rowley, MA: Newberry House Publisher.
- Herrera, L. & Macias, D. (2015). A call for language assessment literacy in education and development of teachers of English. *Colombian Applied Linguistics Journal* 17(2), 302-312.

## List of References

- Hogan-Brun, G., Mar-Molinero, C., and Stevenson, P. (Eds.) (2009). *Discourses on Language and Integration: Critical Perspectives on Language Testing Regimes in Europe*. Amsterdam: John Benjamins.
- Holliday, A. (1994). *Appropriate methodology and social context*. Cambridge: Cambridge University Press.
- Hoyt, K. (2005). *Assessment: Impact on instruction. New vision on Action*. National Assessment Summit, Alexandria.
- Hughes, A. (1989/2003). *Language Testing for Teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (2014). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Huhta, A., Hirvalä, T., and Banerjee, J. (2005). European Survey of Language Testing and Assessment Needs. Part 2: Regional Findings. EALTA. Available online: [http://users.jyu.fi/~huhta/ENLTA2/First\\_page.html](http://users.jyu.fi/~huhta/ENLTA2/First_page.html).
- IBM Corporation. (2017). *IBM SPSS Statistics for Windows, Version 25*. Armonk, NY: IBM Corporation.
- Inbar-Lourie, O. (2008a). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing* 25(3), 385 – 402.
- Inbar-Lourie, O. (2008b). Language assessment culture. In Shohamy, E. and Hornberger, N. (Eds.) *Encyclopedia of Language and Education: Vol. 7, Language Testing and Assessment*. New York: Springer, 285 – 300.
- Inbar-Lourie, O. (2013). Guest Editorial to the special issue on language assessment literacy. *Language Testing*, 30(3) 301-307. <https://doi.org/10.1177/0265532213480126>.
- INEE. (2005). *Políticas y sistemas de evaluación educativa en Mexico: Avances, logros y desafíos. Informe para la Reunión Ministerial del Grupo E-9 (UNESCO)*. Retrived November 2017: <https://www.inee.edu.mx/wp-content/uploads/2019/01/P1D206.pdf>
- Jafarpour, A. (2003). Is the test constructor a facet? *Language Testing*, 20(1), 57-87.
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The international Journal of Research in Teacher Education*, 6(2), pp 26-37.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language

- 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.
- Johnson, D., & Kress, G. (2003). Globalisation, literacy and society: Redesigning pedagogy and assessment. *Assessment in Education: Principles, Policy & Practice*, 10(1), 5-14.
- Jones, N., Saville, N., & Salamoura, A. (2016). *Learning oriented assessment* (Vol. 45). Cambridge University Press.
- Kaiser, G., & Willander, T. (2005). Development of mathematical literacy: Results of an empirical study. *Teaching mathematics and its applications*, 24(2-3), 48-60.
- Kanselaar, G. 2002. *Constructivism and socio-constructivism*.
- Khadijeh, B. & Amir, R. (2015). Importance of Teachers' Assessment Literacy. *International Journal of English Language Education* 3 139. 10.5296/ijele.v3i1.6887.
- Krashen, S. D. (1987). *Principles and practice in second language acquisition*. Prentice-Hall International.
- Krashen, S. D. (1988). *Second language acquisition and second language learning*. Prentice-Hall International.
- Krueger, R. A. & Casey, M. A. (2000). *Focus Groups. A Practical Guide for Applied Research* (3rd Edition). Thousand Oaks, CA: Sage Publications
- Kunnan, A. J. (1998). *Validation in Language Assessment*. London
- Kunnan, A. J. (2009). Politics and legislation in citizenship testing in the United States. *Annual Review of Applied Linguistics* 29, 37 – 48.
- Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests. A Teacher's Book*.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197.
- Lantolf, J. P. (2009). Dynamic Assessment: The Dialectic Integration of Instruction and Assessment. *Language Teaching*, 42, 3, 355 – 368.
- Lantolf, J. P., & Poehner, M. E. (2008a). Dynamic Assessment. In Shohamy, E., and Hornberger, N. H. (Eds.) *Encyclopaedia of Language and Education*, Vol.7, Language Testing and Assessment. New York: Springer, 273 – 284.

## List of References

- Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research. *Language Testing, 28*(3), 367-382.
- Long, M. (1991). Focus on form: A design feature in language teaching methodology. In De Bot, K, Ginsberg, R, and Kramsch, C. (Eds) *Foreign language research in cross-cultural perspective*. Amsterdam: John Benjamins, 39–52.
- Lowenthal, P., and Muth, R. (2008). Constructivism. In E.F. Provenzo, Jr.(Ed), *Encyclopedia of the social and cultural foundations of education*. Thousand Oaks, CA:Sage.
- Lukin, L.E., Bandalos, D.L., Eckhout, T.J. and Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice, 23*(2): 26–32.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Lynch, B. K. (1995). *Language program evaluation*. Cambridge: Cambridge University Press
- Lyons, J. (1981). *Language and linguistics: An introduction*. Cambridge: Cambridge University Press.
- Malone, M. (2008). Training in language assessment. In Shohamy, E. and Hornberger, N. (Eds.) *Encyclopedia of Language and Education: Vol. 7, Language Testing and Assessment*. New York: Springer, 225 – 239.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing, 30*(3), 329–344. <https://doi.org/10.1177/0265532213480129>
- Marso, R. N., & Pigge, F. L. (1993). *Teachers' testing knowledge, skills, and practices*. Lincoln: University of Nebraska
- Martínez Guevara, J., Killian Reyes, M., Audrey Delgado, N., & Brande Martínez, E. (2012). Diseño de planes de estudio de licenciaturas en idioma inglés con el enfoque por competencias: algunas directrices. V Foro Internacional De Especialistas En Enseñanza De Lenguas “Gregorio Torres Quintero.
- Martinez Rio, F. (2013). Dificultades para implementar la evaluación formativa: revisión de literatura. *Perfiles Educativos 35* (139), 128-150.
- Martínez, J., Killian, M., Audrey, N., & Brande, E. (2012). Diseño de planes de estudio de licenciaturas en idioma inglés con el enfoque por competencias: algunas directrices. V Foro Internacional De Especialistas En Enseñanza De Lenguas “Gregorio Torres Quintero.

- Martinez-Rizo, F. (2012). La evaluación formativa del aprendizaje en el aula en la bibliografía en inglés y francés. *Revista Mexicana de Educacion*, 17 (54), 849-875.
- Mavrommatis, Y. (1997). Understanding assessment in the classroom: Phases of the assessment process—the assessment episode. *Assessment in Education: Principles, Policy & Practice*, 4, 381–400.
- McGaw, B. (2008). The role of the OECD in international comparative studies of achievement. *Assessment in Education: Principles, Policy and Practice* 15(3), 223 – 243.
- McLaughlin, B. (1990). Restructuring, *Applied Linguistics* 11 (2) 113-128.
- McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(11).
- McNamara, T. (1996). *Measuring Second Language Performance*. Harlow: Longman/Pearson.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- McNamara, T. (2008). The socio-political and power dimensions of tests. In Shohamy, E. and Hornberger, N. (Eds.) *Encyclopedia of Language and Education; Volume 7: Language Testing and Assessment*. New York: Springer, 415 – 427.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell Publishing.
- McNemar, Q. (1951). The factors in factoring behavior. *Psychometrika*, 16, 353 – 359.
- Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120, 285–296.
- Mertler, C. A. (2002). *Classroom assessment literacy inventory*. (Adapted from the Teacher Assessment Literacy Questionnaire (1993), by Barbara S. Plake & James C. Impara, University of Nebraska-Lincoln, in cooperation with The National Council on Measurement in Education & the W.K. Kellogg Foundation.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12, 101–113.  
<http://dx.doi.org/10.1177/1365480209105575>

## List of References

- 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, April, Montreal, QC, Canada.
- Microsoft Corporation. (2010). Microsoft Excel. Redmond, Washington: Microsoft.
- Miles, M. B. and Huberman, A. M. (1994). *Qualitative Data Analysis*. Second Edition. London: Sage.
- Moore, C. (2017) Current trends in English language teaching in Mexico, in Grounds, P. and Moore, C (eds.) *Higher Education English Language Teaching and Research in Mexico*. Mexico: British Council and University of Southampton: pp: 125-157.
- Moore, M. G. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical Principles of Distance Education* (pp. 22–29). New York: Routledge.
- Muñoz, A. Palacio, M. & Escobar, L. (2012). Teachers' Beliefs about Assessment in an EFL Context in Colombia. *Profile Issues in Teachers' Professional Development*. 14. 143-158.
- Mussawy, S. A. J. (2009). Assessment practices: Student's and teachers' perceptions of classroom assessment.
- Newfields, T. (2006). Teacher development and assessment literacy Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference. (pp. 48-73).
- Nunan, D. (1992). *Collaborative Language learning and Teaching*. Cambridge: Cambridge University Press.
- Nunan, D. (1995). Closing the gap between learning and instruction. *TESOL Quarterly*, 29(1), 133-158.
- O'Loughlin, K. (2006). Learning about second language assessment: insights from a postgraduate student on-line forum. *University of Sydney Papers in TESOL*, 1(1), 71–85.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380.
- O'Malley, J. and Pierce, L. V. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. Boston, MA: Addison-Wesley.
- O'Sullivan, B. (2017). What do language teachers know about assessment? Open Plenary presented at the *Associação Portuguesa Professores Inglêss*. Lisbon, Portugal

- OECD (2010). Acuerdo de cooperación México-OCED para mejorar la calidad de la educación de las escuelas mexicanas. Retrieved from <https://www.oecd.org/education/school/46216786.pdf>
- OECD. (2012). Educational Evaluation and Assessment in Mexico: Strengths, challenges and policy makers. Reviews of Evaluation and Assessment in Education: Mexico.
- O'Sullivan, R. G., & Chalnick, M. K. (1991). Measurement-Related Course Work Requirements for Teacher Certification and Recertification. *Educational Measurement: Issues and Practice*, 10(1), 17.
- Otero, V. K. (2006). Moving beyond the "Get it or don't" conception of formative assessment. *Journal of Teacher Education*, 57, 247–255.
- Oxford English Dictionary (2012). Oxford English Dictionary. Oxford
- Pallant, J. (2013). A step by step guide to data analysis using IBM SPSS.: Survival Manual. United Kingdom: McGrawHill.
- Paterno, J. (2001). Measuring success: A glossary of assessment terms. In Building cathedrals: Compassion for the 21st century. Retrieved July 24, 2003, from <http://www.angelfire.com/wa2/buildingcathedrals/measuringsuccess.html>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402.  
<http://dx.doi.org/10.1177/0265532213480337>
- PISA. (2000). Study. Social Science Research Network: ISA Discussion Paper 739.
- Plake, B. & Impara, J. C. (1993). Assessment Competencies of Teachers: A National Survey. *Educational Measurement: Issues and Practice* 12(4), 10 – 12.
- Plake, B. & Impara, J. C. (1996). Teacher Assessment Literacy: What do teachers know about assessment? In Phye, G. D. (Ed.) *Handbook of*
- Plake, B. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.
- 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.

## List of References

- Popham, W. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265-273. <https://doi.org/10.1080/08878730.2011.605048>
- Popham, W. J. (1991). Appropriateness of Teachers' Test-Preparation Practices. *Educational Measurement: Issues and Practice*, 10: 12-15. doi:[10.1111/j.1745-3992.1991.tb00211.x](https://doi.org/10.1111/j.1745-3992.1991.tb00211.x)
- Popham, W. J. (2004). All about accountability / Why assessment illiteracy is professional suicide. *Educational Leadership*, 62(1), 82-83.
- Popham, W. J. (2006). Needed: A dose of assessment literacy. *Educational Leadership*, 63, 84-85.
- Quilter, S. M., & Gallini, J., K. (2000). Assessment literacy overlooked: A teachers' educator's confession. *The Teachers Educator*, 46(4), 4-11.
- Ramírez Romero, J. L., Pamplón, E. N., & Cota Grijalva, S. (2012). Problemática de la enseñanza del inglés en las primarias públicas de México: una primera lectura cualitativa. *Revista Iberoamericana De Educación*, 60(2), 7. <https://doi.org/10.35362/rie6021321>
- Rasinger, S. M. (2008). *Quantitative Research in Linguistics: An introduction*. United Kingdom: Bloomsbury.
- Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: a learning-focused interpretation. *International Journal of Applied Linguistics* 16, 2, 163 – 188.
- Rea-Dickins, P. (2007). *Classroom-Based Assessment: Possibilities and Pitfalls*. 10.1007/978-0-387-46301-8\_36.
- Rea-Dickins, P. (2008). Classroom-Based Language Assessment. In Shohamy, E., and Hornberger, N. H. (Eds.) *Encyclopedia of Language and Education*, Vol.7, Language Testing and Assessment. New York: Springer, 257 - 271.
- Rea-Dickins, P., & Germaine, K. (1992). *Evaluation*. Oxford: Oxford University Press.
- Reynolds, C. R., Livingstone, R. B., & Willson, V. (2009). *Measurement and assessment in education* 2<sup>nd</sup> ed. United States of America: Pearson
- Richards, J. C. & Nunan, D. (1990). *Second language teacher education*. Cambridge: Cambridge University Press
- Richards, J.C., & Farrel, T. S.C. (2005). *Professional development for language teachers*. Cambridge: Cambridge University Press.

- Rizo, F. M., & Blanco, E. (2010). La evaluación educativa en México: experiencias, avances y desafíos. *Los grandes problemas de México. VII. Educación*, 89-citation\_lastpage.
- Robinson, P., & Ellis, N. C. (2008). An introduction to cognitive linguistics, second language acquisition and language instruction. *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp.3-24). Publisher: Routledge
- Rodriguez, C., & Zabala, J. (2016). Language assessment literacy for test developers: towards an understanding of test constructs. Paper presented at EALTA, Valencia. Spain
- Roux, R. (2018). The future of English Language Teaching in Mexico: Challenges and Opportunities, *Conferencia Regional De Mextesol Capitulo Tamaulipas*, CONALEP, 2 October. Ciudad Victoria: MEXTESOL Tamaulipas Centro
- Roux, R., & Anzures González, E. E.. (2015). Learning Strategies and their Relationship with Academic Achievement in Students of a Private High School. *Actualidades Investigativas en Educación*, 15(1), 324-340.
- Roux, R., & Mendoza-Valladares, J. L. (2014). *Desarrollo profesional Continuo de los docentes: Teoría, investigación y práctica*. El Colegio de Tamaulipas. Mexico
- Rowe, K. J., & Hill, P.W. (1998). Modeling educational effectiveness in classrooms: The use of multilevel structural equations to model students' progress. *Educational Research and Evaluation*, 4(4), 307–347.
- Saefurrohman, S., & Balinas, E. S. (2016). English Teachers Classroom Assessment Practices. *International Journal of Evaluation and Research in Education (IJERE)*, 5(1), 82-92.
- Saunders, M., Lewis, P. & Thornhill, A., (2012). *Research methods for business students*. 6th edn. Pearson Education.
- Savin-Baden, M. & Major, C.H., (2013). *Qualitative Research: The Essential Guide to Theory and Practice*. Routledge.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Australia Language Testing* 30(3) 309–327. <http://dx.doi.org/10.1177/0265532213480128>
- Scovel, T. (1998). *Psycholinguistics*. Oxford: Oxford University Press.

## List of References

- SEP (2012). Acuerdo número 592 por el que se establece la Articulación de la Educación Básica, México. Retrived from <http://basica.sep.gob.mx/dgdc/sitio/pdf/PlanEdu2011.pdf>
- SEP (2011). Plan de estudios 2011. Educación Básica.
- SEP. (2010). *Programa Nacional de Inglés para Educación Básica*. México: Secretaría de Educación Pública.
- Sethusha, M. J. (2014). Communicating assessment results: Teachers' views of recording and repotting classroom assessments. *Mediterranean Journal of Social Sciences*, 5(2) 273-276.
- Sheehan, S. & Munro, S. (2017) Assessment attitudes, practices and needs in ELT teachers. *British Council Report*.
- Shepard, L. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher* 29(7), 4 – 14.
- Shohamy, E. (2001). *The Power of Tests*. Harlow: Longman/Pearson.
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education*, 14 (1) pp. 117-130.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, 57(1), 1-23.
- Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers'assessment literacy. *Journal of Science Teacher Education*, 22(4), 371–391. <http://dx.doi.org/10.1007/s10972-011-9231-6>.
- Sikka, A., Nath, J. L., & Cohen, M. D. (2007). Practicing teachers' beliefs and uses of assessment. *International Journal of Case Method Research & Application*. 19(3), 239-253
- Slade, C., and Möllering, M. (Eds.) (2010). *From Migrant to Citizen: Testing Language, Testing Culture*. Basingstoke, UK: Palgrave Macmillan.
- Soneville, S. (2007). *Fairness in Language Assessment*. Blackwell: USA
- Song, E., & Koh, K. (2010, August). Assessment for learning: Understanding teachers' beliefs and practices. Paper presented at the 36th Annual Conference of the International Association of Educational Assessment (IAEA) on "Assessment for the Future Generations", Bangkok, Thailand.
- Spolsky, B. (1995). *Measured Words*. Oxford: Oxford University Press.

- Stiggins, R. (1991). Assessment Literacy. *The Phi Delta Kappan* 72(7), 534 – 539.
- Stiggins, R. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 238–245.
- Stiggins, R. (1997). *Student Centred Classroom Assessment*. Upper Saddle River: Prentice Hall.
- Stiggins, R. (1999). Are you assessment literate? *The High School Journal*, 6(5), 20-23.
- Stiggins, R. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5-15.
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-65.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 1.
- Sundberg, M. D. (2002). Assessing student learning. *Cell Biology Education*, 1, 11–15.
- Tarone, E. (2005). Speaking in a second language. In E. Hinkel (ed.), *Handbook of Research in Second Language Teaching and Learning*, 485-502, NJ: Lawrence Erlbaum Associates.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics* (29), 21 – 36.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412
- Thomas, G. (2017). *How to do your research project: A guide for students*. London: SAGE Publications.
- Thomas, J., Allman, C., & Beech, M. (2004). *Assessment for the diverse classroom: A handbook for teachers*. Tallahassee, FL.
- Thornbury, S. (2006). *An A to Z of ELT*. Oxford: Macmillan.
- Thorndike, R. L., & Hagen, E. (1961). *Measurement and evaluation in psychology and education* (2nd ed.). Oxford, England: Wiley.
- Tolmie, A., Muijs, D., & McAteer, E., (2011). *Quantitative methods in educational and social research using SPSS*. McGraw-Hill Education (UK).

## List of References

- Tsagari, D. (2004). Is there life beyond language testing? An introduction to alternative language assessment.
- Tsagari, D. (2009). *The complexity of test washback: An empirical study* (Vol. 15). Peter Lang.
- Tsagari, D., & Vogt, K. (2017). Assessment Literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects. *Papers in Language Testing and Assessment*. 6. 41-63.
- Universidad Autónoma de Tamaulipas. (2008). Plan de Estudios Licenciatura en Lingüística Aplicada. Cd. Victoria, Tamaulipas.
- Universidad Autónoma de Tamaulipas. (November 2011) Acuerdo Rectoral REC/002/XI/11. Ciudad Victoria, Tamaulipas.
- Upton, G., & Cook, I. (1996) *Understandings statistics*. Oxford: Oxford University Press.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of an European study. *Language Assessment Quarterly*, 11(4), 374-402
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30, 749–770.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* Cambridge, Mass. Harvard University Press.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge: Cambridge University press.
- Wall, D. (2012). Washback. In Fulcher, G. and Davidson, F. (Eds.) *The Routledge Handbook of Language Testing*. London and New York: Routledge.
- Wallace, M. J. (1991). *Training foreign language teachers: A reflective approach*. Cambridge: Cambridge University Press.
- Walters, F. S. (2010) Cultivating Assessment Literacy: Standards Evaluation through Language-Test Specification Reverse Engineering. *Language Assessment Quarterly* 7(4), 317 – 342.
- Ward, A., & Murray-Ward, M. (1999). *Assessment in the classroom*. Wadsworth Publishing Company.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave: London.

- Widdowson, H. G. (1996). *Linguistics*. Oxford: Oxford University Press
- William, A. (2006). The hope and practice of teaching. *Journal of Teacher Education* 3(57), 269-277.
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241-256.
- Wright, S. P., Horn, P., & Sanders, W. L. (1997) Teacher and classroom contexts effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57–67.
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.
- Yule, G. (1996). *Pragmatics*. Oxford: Oxford University Press
- Zhang, Z., & Burry-Stock, J. A. (1997). Assessment practices inventory: a multivariate analysis of teachers perceived assessment competency. *Paper presented at the Annual Meeting of the Nacional Council on Measurement in Education*, March 25-27, Chicago, Illinois.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), pp.323-342.