UNIVERSITY OF SOUTHAMPTON

FACULTY OF HUMANITIES

Modern Languages

The Impact of Assessment Training on English as a Foreign Language University Professors´ Classroom Writing Assessment: Reported Practice and Perceptions

by

Elsa Fernanda González

Thesis for the degree of Doctor of Philosophy

March, 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF HUMANITIES

Modern Languages

Thesis for the degree of Doctor of Philosophy

THE IMPACT OF ASSESSMENT TRAINING ON ENGLISH AS A FOREIGN LANGUAGE UNIVERSITY PROFESSORS´ CLASSROOM WRITING ASSESSMENT: REPORTED PRACTICE AND PERCEPTIONS

Elsa Fernanda González

This study analysed the impact that two sessions of writing assessment training had on English as a Foreign Language (EFL) Mexican university teachers. It focuses on three main areas of impact, a) teachers´ reported classroom assessment of students´ writing skills, b) teachers´, language program managers´ and students´ perceptions towards writing assessment as well as assessment training and c) the changes that training may have encouraged in teachers´ analytic and holistic scoring.

Forty-eight EFL university teachers and four EFL program managers took part in the initial stage of the study, which included, a participant background questionnaire, the first training session, a pre-training interview and an initial round of analytic and holistic assessment of five opinion essay samples. Additionally, four groups of EFL students took part in a pre-training and post-training focus group interview.

From these forty-eight teachers, eleven continued on to the second phase of the study, by participating in a second training session, a post-training interview, answering a post-training online questionnaire and scoring the same written samples analytically and holistically once more. Participants were tracked for a period of 12 months to analyse the changes that assessment training could have encouraged. Data obtained from the participant background questionnaire, the scores to the five opinion essays, and the online post-training questionnaire were collected and analysed quantitatively while the semi-structured interviews and the student focus groups were conducted and examined under a qualitative approach.

Data suggested that teacher participants and language managers considered the training useful, practical and objective for their future assessment practice. However, the impact of the sessions on classroom assessment practice was quite shallow. Instead, more impact was found in teachers´ reflective processes and self-awareness of themselves as EFL teachers and assessors. A Writing Assessment Training Impact Categorization is proposed so as to classify this impact. It is believed that results may have implications for EFL assessment, language program management and teacher training. Further discussion of results and implications of these for the EFL context in Mexico is provided.

Table of Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, ELSA FERNANDA

GONZÁLEZ…….

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.

The Impact of Assessment Training on English as a Foreign Language University Professors' Classroom Writing Assessment: Reported Practice and Perceptions

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. Parts of this work have been published as:

Gonzalez, 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 City: British Council Mexico, 73-100.

Gonzalez, E.F. (2017) EFL university teachers´ perceptions of writing assessment training, EDUCIENCIA, (5), 11-22.

Signed:

Date: March 08, 2018

Acknowledgements

A special thought of appreciation goes to my main supervisor Dr. Ying Zheng, for her patience, guidance and words of encouragement throughout this journey. This PhD, has been a learning process that has allowed me to grow as a person, as a professional and as a researcher. Thank you, Ying for being a role model in each of these aspects and many more.

My deepest thought of gratitude to my children, Martín Jr., Mateo and my husband, Martín for being the engine of my life, and for sharing the many sacrifices that this journey has implied. To my mother Carolina, my father Arnoldo, and my siblings Guillermo and Franco for always believing in me and my dreams. Thank you to Leticia, Jorge (+), and Claudia for helping out when no one else could. A PhD is not possible without the support of family.

I thank from the bottom of my heart the teacher participants of this study, my enthusiastic students enrolled in the BA in Applied Linguistics at the Universidad Autónoma de Tamaulipas, and my PhD cohort peers for their constant and unconditional support. Their words of wisdom and of comfort in difficult times, gave me strength to move forward.

Finally, I would like to profoundly acknowledge the financial and academic support of my University, the Universidad Autónoma de Tamaulipas, the Mexican PRODEP Program (Teacher Professional Development Program), The British Council Mexico and the British Council Assessment Research Award Grants Committee for recognizing my work and believing in this research project.

Definitions and Abbreviations

ANOVA: Analysis of Variance

CEFR: Common European Framework of Reference for Languages

EFL: English as a Foreign Language

ELT: English Language Teaching

ESL: English as a Second Language

ETS: English Testing Service

EXAVER: Exam of the University of Veracruz

EXIT: English Test of Tamaulipas

FL: Foreign Language

L1: First Language

L2: Second Language

TOEIC: Test of English for International Communication

TOEFL: Test of English as a Foreign Language

WAT: Writing Assessment Training

WATIC: Writing Assessment Training Impact Categorization

Chapter 1: Introduction

The assessment of students´ writing performance is a complex activity that teachers are required to do in their regular teaching practice. In the Mexican English as a Foreign Language (EFL) context, as in many other parts of the world, language instructors need to select an assessment method that corresponds to their assessment purposes, develop the assessment tool to use in the classroom, administer the tool, score the tool, interpret the score, make appropriate decisions, communicate the results to administrative offices and finally be aware of the consequences that assessment may bring (Crusan, 2014; Cumming, 1990; Taylor, 2009; Fulcher, 2012; Stoynoff and Coombe, 2012; Weigle, 2007, Scarino, 2013).

EFL teachers may not have the necessary theoretical and practical skills to assess their students. On the other hand, fair assessment of writing skills needs to consider the local practices and purposes of people involved in the writer´s process, the teaching of EFL writing and the assessment of the skill (Pearson, 2004). This sets forward a difficult context to cope with in the field of classroom assessment due to the fact that, the assessment context is a determining factor when seeking valid and reliable assessment of writing. Additionally, the assessment of writing will always be subject to human judgment therefore providing fair and accurate scores to students´ writing may be quite difficult (Pearson, 2004) to accomplish. However, researchers (Weigle, 1994, 1998, 2007) believe that assessment training may be a tool that can be used to lessen writing assessment difficulty and increase score reliability.

Therefore, this project examines the impact that writing assessment training has on the reported teaching and assessment practices of forty-eight Mexican University EFL teachers and four EFL program managers. It approaches this purpose from three major dimensions: a) teachers´ reported writing assessment practice in the EFL classroom b) teachers´, language managers' and students' perceptions of the assessment of writing and their teachers' assessment literacy and c) teachers´ use of scoring tools to score students´ opinion essay samples. This study also tries to analyse the usefulness of assessment training to promote the teaching and assessment of writing in the Mexican EFL classroom.

This chapter provides an outline of the project and describes the EFL context in the north-eastern part of Mexico as the research context, the background to the study´s development, the need for it in the context of foreign language assessment, its rationale, its significance and finally the organization of this thesis is explained in section 1.5.

1.1 The Research Context: EFL in the State of Tamaulipas, Mexico

In our actual globalized world, the ability to communicate in more than one language is a skill that is highly valued in Mexican students during their undergraduate studies and when they enter the professional world (Universidad Autonóma de Tamaulipas, 2008). Specifically, in the north-eastern region of the country, the state government of Tamaulipas has tried to provide globalized education to students by teaching English in public primary and preschool school curriculums. In 1999, the State Department of Education of Tamaulipas implemented The English Language Program in Primary Schools (Coordinación de Inglés en Educación Básica, 2015) and was piloted in fourth grade of some public elementary schools. Over the years, the program expanded and by 2005 fourth, fifth and sixth grades were included in the program. By 2012, the English program

became obligatory for pre-schoolers in the state and since 2013 the program had been piloted in secondary levels in some public schools.

In the tertiary level, students are provided with English lessons in at least six semesters (private universities) or three semesters of their undergraduate studies (public universities). At the end of their programs, students are expected to provide proof of their English language proficiency as a requirement to obtain their degree diplomas. However, the English language programs that universities follow are autonomous and do not follow a specific standardized curriculum. Some private universities require their students to obtain 500 points or more on the Institutional TOEFL Test while public universities require 450 points or more. The specific number of points required depends on the specific institutional policies and the language program characteristics. Students enrolled in the public institution under analysis in this study are also given the option of presenting the EXIT exam (English Test of Tamaulipas). The EXIT exam is a locally generated exam developed by experienced English language teachers in service at the Centre for Languages and Applied Linguistics located in capital of the state of Tamaulipas: Victoria.

This scenario suggests that the relationship and correspondence among classroom assessment and large-scale testing is crucial and that EFL university teachers need to have the skills to assess their students in a valid and reliable way (Weigle, 2002, 2007). This would help avoid negative washback (Hamp-Lyons, 1990) on students´ academic and personal lives. In other words, it is necessary for classroom assessment of writing, and of other skills, to be linked to large-scale testing so that students have greater opportunities of obtaining satisfactory scores to obtain their undergraduate degrees.

1.2 Background of the study

With the purpose of exploring the EFL context in the region and establishing the background to this project, a preliminary exploratory study was carried out in September of 2013 in which twenty-five EFL university teachers answered a thirteen open and closed-question questionnaire. Eleven of the twenty-five participants were males while fourteen were females, their ages ranged from twenty to sixty years old. They all were part of the EFL teaching staff at eight different schools part of different public universities in the city of Victoria. Data resulting from quantitative analysis of closed-ended questions in combination with qualitative analysis of open-ended questions suggested that in terms of writing assessment, teachers who were part of the same school did not follow specific assessment standards. Teachers assessed without following a specific procedure or specific purpose. Although it is difficult for specific standards to be set among different universities due to their particular characteristics, I consider it is important to set assessment standards within the same institution. This could make the assessment process more reliable and valid.

Secondly, some teachers stated that they did not use a specific standardized assessment tool, but instead modified criteria depending on the units´ content, the learning objectives, and the teaching purposes. Therefore, evaluating students´ texts with a distinct scoring tool each time students wrote. Although assessment standards need to be set according to writers´ context and the program's learning purposes, I consider that using a distinct tool each time a text is evaluated gives the student a sense of insecurity towards his work and diminishes the validity and reliability of the process and therefore of the score.

Finally, other teachers assured they used a scoring rubric and described it as a set of symbols that allowed their students to spot their flaws among their writing and reflect on the possible improvement of flaws. It was deduced from the teachers´ answers that they were confusing the concept of `error correction code´ with that of the `scoring rubric´. This misunderstanding of information could lead to unreliable assessment processes among EFL teachers within Mexican higher education institutions.

Keeping these issues present and the research context in Victoria, Tamaulipas it was believed that the majority of the potential teacher participants of this research project were inexperienced in terms of writing assessment and that the institutions in which they worked did not provide them with the necessary assessment training. However, as an EFL teacher, participants are required by their administrations to assess students´ language skills on a regular basis. Thus, it is considered that, this study may enlighten the path to fulfil these needs, which are further explained in the following section.

1.3 The need for the study

The potential research context paints a difficult and subjective picture: assessment standards are not followed while assessment tools are misused due to the lack of training. In other scenarios, teachers are avoiding the assessment of writing in their classrooms due to the lack of assessment literacy and a sense of uncertainty in their abilities to teach and assess writing. It is my belief that Mexican university teachers in the north-eastern part of the country should experience appropriate training prior to writing assessment. Therefore, it was considered that the need for this study was to satisfy the professional needs of Mexican EFL university teachers by providing writing assessment literacy that can allow them to give more reliable classroom assessment to their language students. The study can

fulfil the need of university language programs and institutions in the region with the necessary data and input that can aid them in establishing assessment standards. This study may also provide the necessary information for stakeholders to consider the importance of writing in a language curriculum, its standardized assessment and the role of teacher assessment literacy in language classroom assessment. Finally, the present study seeks to provide assessment stakeholders with the identification of the possible impact of assessment training so as to identify the type of outcome they wish to produce in their teaching staff when providing training.

Research in writing assessment has explored various aspects in second language contexts, and large-scale testing. Assessment literacy, on the other hand, has been approached to understand the issues teachers encounter in their assessment contexts as well as explain the needs and knowledge that language teachers consider they possess. The following section provides a brief overview of the research that has focused on the assessment of writing and assessment literacy for language teachers.

1.4 Rationale

Researchers have approached writing assessment from different perspectives throughout the years. In the language testing area, studies led by Elder *et al*. (2005, 2007); Weigle (1994, 1998); Contreras, González, and Urias (2009) and Knoch (2011) focus on how rater training has a role in the score provided to the written text in large-scale testing contexts. Other experts such as Barkaoui (2007) and Knoch (2009) have set out to analyse the impact that the use of a specific type of rubric has on writing assessment. Barkaoui (2011), Esfandiari and Myford (2013), Lim (2011) and Wiseman (2012) followed a different purpose and studied the impact that rater background has on the score provided to a text,

scoring procedures followed by the rater and the rating behaviour that a rater portrays. In the Mexican EFL context, rubrics, or scoring tools, are not only used in a standardized-testing context. They are used by teachers as tools to provide feedback to their students about their writing and as guides in the assessment process that is carried out regularly throughout a school period in an institution. Therefore, research into how teachers use these scoring tools seems very useful in this context.

In the field of classroom assessment, research has focused on the assessment of language in EFL/ESL classrooms in contexts such as Canada, USA, China, Iran and Israel (Cheng and Wang, 2007; Cheng, *et al.* 2008; Shohamy *et al*., 2008; Inbar-Lourie and Donitsa-Schmidt, 2009; Ketabi and Ketabi, 2014) in which the main purposes have been to understand what teachers do in the classroom to assess language abilities or to provide a comparison of assessment practices in different parts of the world. Research still needs to consider the Mexican approach to EFL assessment, specifically towards writing skills. Therefore, this study seeks to contribute to the field of language assessment by providing some insight into how teachers tackle and perceive writing assessment in their Mexican EFL classrooms.

In the field of assessment literacy for teachers of elementary levels in the United States, Stiggins (1995), Metler (2003) and Mertler and Campbell (2005) defined the concept of assessment literacy and emphasized the importance of providing teachers sufficient academic preparation so they could assess their students adequately. Mertler and Campbell (2005) presented an Assessment Literacy Inventory that was developed with the purpose of measuring the assessment literacy of teachers according to the Standards for Teacher Competence in the Educational Assessment of Students. In Mexico, studies have been

carried out with English teachers working in the National Education System in secondary school levels (Roux and Valladares, 2014) in which their perception of teacher development programs and their commitment to continuous professional development efforts was surveyed. Teachers of this study identified evaluation and assessment as one of the aspects in which they had the least academic background.

In foreign language (FL) contexts, assessment literacy has been analysed in relation to teachers´ perceptions of training courses (Nier, Donnovan and Malone, 2013; Fulcher, 2012; Jeong, 2013) and the needs that they consider should be covered in an assessment training course. Nier, Donnovan and Malone (2013) analysed online training and its usefulness to language teachers. They concluded that most of the FL teacher participants considered the online training useful for their future assessment practice but more examples and samples were needed to further understand the assessment process. Jeong (2013) presents a study in which Language Assessment Courses (LAC) were provided to teachers and concluded that the ultimate outcome of assessment teacher training courses will largely depend on the academic background and the personality of the instructor even if the structure is similar or the same. Fulcher (2012) examined the assessment training needs of language teachers in Europe with the purpose of producing materials to use for foreign language teacher training programs. The researcher describes his results and explains that it is necessary to understand the role that testing and assessment have in today's society in order to provide teachers with tools that can allow them to understand the principles and the essence of classroom assessment.

Research in the field of assessment literacy and writing assessment has yet to clarify the level of impact that training produces in instructors´ actual writing assessment practice

particularly in the EFL classroom. It is my belief that although research has examined assessment training and its impact on a variety of areas; the specific longitudinal changes that training may cause in experienced and non-experienced teachers´ small-scale classroom assessment, on their perceptions as teachers, is still underexplored. Additionally, I consider that insight from the different stakeholders involved in the assessment procedure such as experienced and novice teachers, students and language program managers could further explain the changes generated in the assessment practice of teacher participants once assessment training has been delivered. Therefore, this study has the purpose of analysing the impact that assessment training has on EFL teachers´ reported assessment of students´ writing skills. It mainly focuses on changes in a) teachers´ regular writing assessment procedures in the classroom, b) teachers´ use of assessment tools to assess a written text and c) the perceptions that teachers, language program managers and students have of writing assessment and writing assessment training.

The significance of the results of this study lay in the possibility of raising awareness among EFL teachers and heads of language departments of the importance of providing teacher assessment literacy as a means of seeking valid and consistent EFL writing assessment. Therefore, the results of this study could emphasize the importance that writing assessment standards and teacher training have to a language program in our universities. It will hopefully persuade language managers to give more importance to the professional development of their teaching staff and the establishment of context-specific assessment standards. The following section describes the organization of this document to provide an overview of the study.

1.5 Organization of the thesis

The thesis is organized in seven different chapters. Chapter One focuses on providing the reader with a general background to the study. It includes an overview of the scenario presented in the Teaching of English as a Foreign Language in Tamaulipas Mexico (research context). It then goes on to describe the need for this research project and its significance finalizing with the rationale behind the research purposes.

Chapter Two focuses on the concepts of writing, its skills, and writing process models suggested by Hayes since 1981 to 2012. It tries to describe writing as a social skill in which the audience and the writer interact. This Chapter moves on to describe language assessment and the assessment of writing in the EFL classroom. It intends to provide a description of the literature review that provides a theoretical background to the study. The chapter overly discusses the concept and importance of writing assessment for EFL classrooms. It also describes the concepts of analytic and holistic scoring tools and important factors to consider when assessing writing.

Chapter Three goes on to explain the concept of assessment literacy and the importance it has in the EFL context. The Chapter tries to contextualize the practice of assessment literacy in Mexico and how researchers suggest this literacy be encouraged in language teachers.

Chapter Four describes the methodology followed throughout this study. It begins by recalling the purpose of the project and then moving on to the description of the method of this study. It explains the research context, the participants involved, and the data collection instruments. Then, the data collection procedures are depicted in phases and the

chronological order in which they were conducted. This Chapter also approaches the assessment training sessions and the process carried out to adapt and pilot the holistic and analytic rubric. Finally, this fourth Chapter focuses on explaining the process followed for the analysis of information.

Chapter Five carries on drawing attention to the results obtained from the background questionnaire, the interview to participants´ of the study and the online questionnaire answered by these participants. It describes the teacher participants' and the stakeholders' perspectives as well as the difficulties faced in relation to the teaching and assessment of writing. Then, it goes on to point out the results obtained from the quantitative analysis of the holistic and analytic scores provided to five written samples of an opinion essay prior and post to the training sessions. It details the results obtained of analysing how teachers´ personal and academic background had an impact on their use of scoring tools to score the opinion essays.

Chapter Six provides my insight of the data obtained to answer the five research questions (RQs) that lead this study as well as my final perspective of the results obtained in this project. It also describes the Writing Assessment Training Impact Categorization (WATIC, Figure 7) as one of the main findings of this study. The discussion is organized in accordance to the order of the RQs and is nourished by research studies carried out by other experts that can exemplify the results obtained in this project. This Chapter also points out to the reader the researchers´ point of view in regard to the results obtained.

Chapter Seven provides some concluding remarks that arouse from the results obtained in this study as well as the limitations that this study dealt with. It exemplifies the

contributions that I believe this research project provides to the language assessment and assessment literacy fields. It then moves on to point out possible research ideas that are born from the limitations of this study and finalizes with the implications that the results of this project may involve for the EFL classroom, curriculum and teacher assessment literacy.

Chapter 2: The Nature of Written Language and Language Assessment

The following chapter has the purpose of presenting the supporting literature that was considered for the development of this project. First a description of the nature of writing, the writing process and writing genres is provided. Then, an overview of language assessment, writing assessment and the use of scoring rubrics is given. Finally, the assessment of writing the issues it may entails are explained.

2.1 Towards a model of L1 and L2 writing

Drawing upon a first language (L1) linguistic perspective, writing was a secondary system of study immersed in a context that gave more importance to phonological and spoken systems (Daniels, 2001) for nearly a century. It was differentiated from language and treated separately as stated by Grabe and Kaplan (1996, p.3),

> Writing is a rather recent invention, historically speaking. Unlike spoken language- coterminous with the history of the species- written language has a documented history of little more than 6000 years. And while it is generally accepted by linguists that certain aspects of spoken language may be biologically determined, the same cannot be said about writing.

It was treated separately for several reasons. Writing as a linguistic system is not innate, in other words, humans need to be instructed how to write. It is acquired and produced consciously rather than unconsciously like other language systems such as the phonological system used in spoken speech. Additionally, writing systems tend to follow a specific tradition. They do not follow an evolving pattern such as language. Most importantly, linguists considered that the disappearing essence of spoken production versus the physical existence of a written system called for it to be analysed differently and apart

of language systems (Grabe and Kaplan, 1996). Some researchers suggested it be analysed as a linguistic system, that under the umbrella of semiotics, conserves and allows language to transcend (Daniels, 2001).

It was not until the 70´s and 80´s when writing began to be further analysed to provide cognitive and psychological models to further understand its process as a language system. Flower and Hayes´ (1981) model was the first attempt to describe the cognitive processes that L1 writers experience in their text production. The model (Figure 1) they proposed considered writing a system that is influenced by three crucial factors `the task environment, the writer's long-term memory, and the writing processes´ (Ibid, p.369).

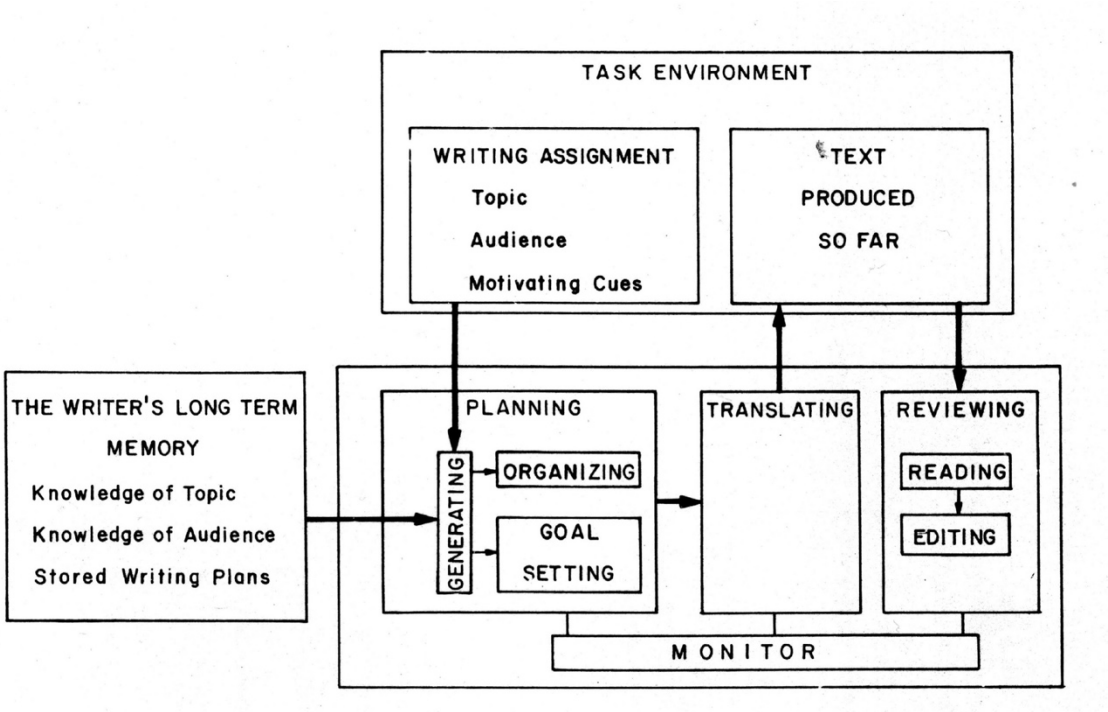

Figure 1: Flower and Hayes´ (1981, p.370) Model of Writing

Since then, Hayes (1996, 2012) has updated the Model twice. As shown in Figure 2, the updated Model gave emphasis not only to long-term memory but also to a working

memory that is available to writers in any phase of their text production (Hayes, 2006). This working memory allows the writer to activate his experiential schema to produce a text through different stages. Additionally, this Model included affective factors of the writer: motivation, beliefs, attitudes, goals, predispositions interacting with the social environment and physical environment the task is written in. Therefore, this model acknowledged the importance of these factors not only for the final product but also for the different stages a writer undergoes. As Deane *et al*. (2008, p.5) specify,

> In this revised model, Hayes (1996) sought to identify how various aspects of human cognitive capacity interact with these tasks, distinguishing the roles of long-term memory, short term memory, and motivation or affect. The Hayes (1996) model is specific about the contents of long-term memory, distinguishing among task schemas, topic knowledge, audience knowledge, linguistic knowledge, and genre knowledge. Similarly, Hayes (1996) specified how different aspects of working memory (e.g., phonological memory and visuospatial memory) are brought to bear in the cognitive processes of writing.

Finally, in the most recent Model by Hayes (2012) the writing process is considered to be composed of three levels, which are connected among each other: Control Level, Process Level and the Resource Level (Figure 3). Features such as the working and long-term memory are kept in the Resource Level. The Process Level is formed by the task environment and the actual writing process. Agents such as collaborators, critics, task materials, transcribing technology and written drafts are involved in this level. This newest model acknowledges the importance of the roles the writer may take during text production such as a translator, proposer, evaluator and transcriber.

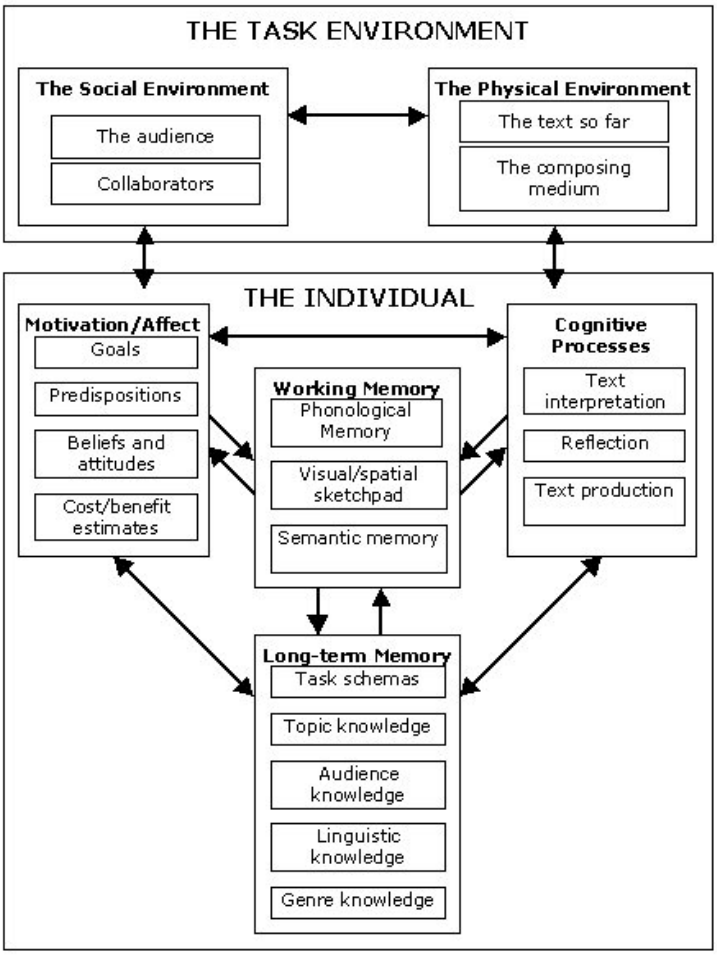

Figure 2. Framework for understanding cognition and affect in writing (Hayes, 1996)

Specifically, Hayes (2012) draws attention to the transcriber role for he considered, as other authors (De La Paz and Graham, 1995; Jones and Christensen, 1999; Christensen, 2004 cited in Hayes, 2012) considered, that transcription activities such as spelling, typing practice or handwriting activities had a role in the quality of the written text therefore suggesting a cognitive process involved.

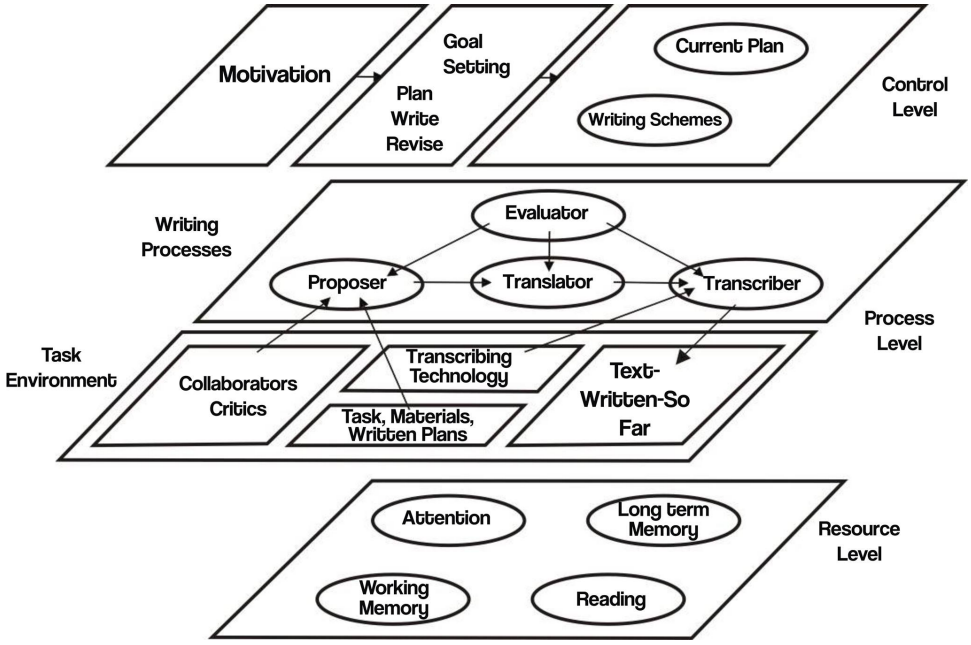

Figure 3 Updated Model of Hayes´ Writing Process (2012)

When comparing the Models depicted in Figure 1 (pg.14) and Figure 2 (pg.16), it can be noticed that the writers´ individual characteristics and the role of these in the production of texts are reflected in greater depth in Model 2 (Figure 2). The writers´ affective factors, cognitive processes and long-term knowledge such as the linguistic schema and knowledge of the audience are also portrayed. Another factor which I consider is worth noticing, is the inclusion of the writers´ goals in Model 2 while Model 1 includes only the process of goal setting rather that the goal itself. This element as well as the element of the `audience´ is reflected in both Models. However, in the first Model (Flower and Hayes, 1981) the audience is considered only as part of `The Task Environment´ while in the second Model (Hayes, 1996) it is also considered part of the students´ long-term knowledge. The inclusion of this element at this level may allow the consideration of the audiences´ potential expectations of the text.

Finally, the Model included in Figure 2 includes the element of `Genre Knowledge´ of the writer but not in the ` The Task Environment´ section of the figure. It is my belief that to write a specific genre of writing it is necessary to have the cognitive knowledge of its specific characteristics but also the involvement of a genre in a specific context (The Task Environment) in the production of a text. Therefore, I would propose the inclusion of Genre Knowledge in both the constructs of ` The Task Environment´ and `The Individual´.

The third Model (Hayes, 2012), as portrayed in Figure 3, attempts to portray the different levels in which the writer interacts with her/his own processes, the resources available to produce the text and the factors that are controlled during the writing process. The role of `The Task Environment´ is also recognized in this Model such as in Model 1 and 2 with additional elements which include technology, materials, the text written so far, and collaborators as critics. However, a crucial element such as the intended audience and its expectations have not been considered in this update. Additionally, the factors of the Writing Genre Knowledge and the production of a specific genre in a given context are not considered. It is my belief that, without a clear idea of the specific characteristics of a text, the intended audience and the involved context, the produced text will not transmit or maintain a main idea.

It is worthwhile mentioning that these models portray the possible cognitive processes and elements that an L1 writer may experience but for L2 or FL learners these models may need further adaptation. For instance, students who are learning an L2 or FL have different linguistic knowledge than native speakers. Therefore, I believe linguistic knowledge should be included to understand how the learner uses this knowledge to produce a text in addition to other elements such as intended audience, audience expectations, (for instance

what teachers expect to read in student´ texts), and finally students´ affective traits (such as motivation, anxiety, self-awareness). Therefore, I would reconsider the role of these models in the EFL classroom and their true portrayal of the actual writing activity in a FL classroom. I believe it would be very difficult for a language teacher to encourage her students to experience every stage of this process due to time constraints or other factors involved thus my argument in favour of the creation of FL writing models that consider the role of context as an important factor.

Although these previously described models have contributed to teachers´ and other researchers´ understanding of the role of cognitive, social and affective processes present in the creation of a text written in L1, research has yet to clarify the processes that L2 and FL learners experience during their composition experiences (Polio and Williams, 2011). Studies such as Zamel´s (1983), who focused on L2 learners of English, and more recently Sasaki´s (2000), who analysed EFL writers, have attempted to understand the processes that novice and more-skilled writers of English experience while producing a text.

Other researchers such as Wang and Wen (2002) provide a description of these processes, there still exists a lack of connected theory that could explain the processes of EFL student writers that could lead to the creation of an EFL/ESL writing model. However, it seems only logical to discuss the models proposed in L1 contexts as a means to understand the nature of writing, its processes and genres in EFL/ESL contexts. The following sections attempt to discuss firstly the essence of writing, the distinct and most common genres taught in the Mexican EFL context and finally the needs of EFL students.

2.2 The nature of written language

Several authors have attempted to explain what writing is and how humans represent it. A linguistic point of view to writing is provided by Daniels (2001, p. 68) who considers writing to be a system of permanent markers that represent an utterance. Byrne (1991) and Brown (2007) provide a more communicative explanation to writing by considering it a system in which an alphabet or a set of symbols is comprised together to create meaning and communicate with others. Byrne (1991) adds that beyond the production of symbols, the existence of an established order to arrange symbols and sentences is necessary.

Similarly, Ferris and Hedgecock (2014, p.5) conceptualize writing as a type of communicative `…system that combines semiotic, communicative, cognitive, and creative functions´ therefore giving importance to other aspects of the ability such as the meaning that the author wishes to convey, the mental skills needed to produce a text and the creativity that a writer can integrate to a composition. In a deeper analysis of these factors that comprise writing, Hyland (2015) considers that the analysis of the text on its own is not enough. It is also necessary to account for the social role that the writer and the reader have in the composing process. When focusing on writing as a text, Hyland (2015) provides an understanding of it as an object of language or as an object of discourse. From the perspective of the writer, writing is considered an exemplification of words, structures and clauses that follow a specific order and its use can provide a scope of the writer´s mastery of these grammar rules, the prior focuses on written text as an example of language in action to convey meaning in a social context. In this dimension, the writer has a specific intention and it is through the discourse produced that these intentions are accomplished (Ibid, p.6).

Another way of analysing the nature of writing is by describing the writers´ purpose when for instance used to identify, communicate, call to action, to recall, to satisfy a requirement, to introspect, to create (including a combination of information or the creation of knowledge) (Grabe and Kaplan, 1996). In other words, writing from a writer´s point of view may include a more text-level intention of composing a text or surface-text level intention that may not include a composing process. Hyland (2015) agrees to this view and explains that a writer may intend to write as a source of self-discovery and cognitive maturity as a result of the written text.

In the Mexican EFL context, writing is considered a skill that needs to be learned to acquire communicative language competencies (Universidad Autónoma de Tamaulipas, 2011), rather than a process from which students and teachers can learn. In my experience as an EFL teacher and considering the programs that teachers and students need to cope with, writing is a skill that, portrays students´ abilities to combine their linguistic knowledge, their knowledge of specific writing conventions, and the specific contextual traits in which they write to communicate a specific idea or point of view. It is a window that may also allow teacher-student understanding since writing allows for detailed expressions of thought. However, in the reality that EFL teachers face, this skill is sometimes excluded from their teaching in the classroom because of different issues that may include fear, rejection, or lack of time. These and other variables are further described in Chapter 5.

Student language learners are required by the educational system to gradually move from a surface-level text product in low proficiency level (such as form filling, formal and informal email writing, among others) English courses to a more profound text-level stage

of their writing in which the student must create meaning through journal entries, poems or stories considering the audience the text is created for. Considering this broad context, Mexican EFL students have different language and professional needs that impact their writing development as a language skill as well as the teaching and assessment of writing in the EFL classroom. Therefore, the following section provides a more in depth description of this context.

2.3 Teaching Writing in the EFL Context

Students immersed in an EFL context learn the language in a country whose official language is not English; such as the case of Mexico, China or Japan (Grabe and Kaplan, 1996; Polio and Williams, 2011). The English as a Second Language (ESL) student learns English in a country that considers it the language of communication (Ferris and Hedgcock, 2014). Although, some theoretical and methodological aspects to the teaching and learning of the language may be similar in both contexts there are other aspects that differ greatly due to the vast number of differences between both contexts.

The needs of EFL students in the Mexican context are very similar to those outlined by Grabe and Kaplan (1996, p. 25), they `…will need English writing skills from simple paragraph writing and summary skills to the ability to write essays and professional articles depending on students´ educational levels, academic majors and institutional demands´. As in countries like China, EFL writing instruction at University level is seen as part of a holistic approach to the development of English language skills. In other words, writing is not taught as a separate course that dedicates its total number of hours exclusively to writing abilities. Instead EFL teachers need to balance their classroom time among the four language skills and sub skills their language programs specify. On the other hand, tertiary

EFL teaching and learning in Mexico has been greatly influenced by a `teaching for tests´ culture that has a great impact on students´ future academic and professional lives. In other words,

> … in spite of many teachers´ awareness of more labor-intensive approaches to writing instruction, including genre and process pedagogies, the realities of large classes, students´ relatively low proficiency, overworked and underpaid teachers and lack of teacher preparation have forced most teachers to teach to these tests…which may not assess writing directly, instead testing sentence level knowledge and the ability to reproduce models (Polio and Williams, 2011, p.494).

This present situation in the Mexican EFL classroom and other classrooms around the world lead to the question of how can the assessment of writing be further used as a trigger to raise awareness of the importance of developing writing in the English language classroom? Can teacher assessment training lead to more valid and consistent classroom assessment? These and other questions are approached in this project. However, before moving on to the description of this subjective situation it seems relevant to discuss the basic components of language assessment.

2.4 Language Assessment and its importance to Language Development

Language assessment is used widely around the world for several different purposes. Usually, decision-making based on the information collected throughout the assessment process has vital consequences for stakeholders, institutions, language managers, language teachers and undoubtedly for students. As pointed out by Bachman and Palmer (2010, p.22), `The primary use of any language assessment is to collect information to make decisions. These decisions have important consequences for stakeholders, the individuals and programs in the educational and societal setting in which assessment takes place´. This following section describes the concepts of assessment and evaluation while it discusses the importance of providing valid and reliable assessment to language students.

Among the field of language teaching, there seems to exist different concepts of assessment. For instance, O´Malley and Pierce (1996, p.1-2) emphasize the concept of authentic assessment as an alternative method of classroom assessment by pointing out that,

> Alternative assessment consists of any method of finding out what a student knows or can do that is intended to show growth and inform instruction, and is an alternative to traditional forms of testing namely, multiple-choice tests. Alternative assessment is by definition criterion-referenced and is typically authentic because it is based on activities that represent classroom and real-life settings.

With this concept, the authors emphasize the importance of existing connections among classroom instruction, language curricula, and language assessment. To differentiate assessment from testing, Brown (2007) considers assessment to be a process that involves much more than testing. While testing is a specified method to measure a person's ability in an established topic, assessment is an `ongoing process´ (p.445) in which the language teacher is constantly observing and judging students´ performance or the teaching practice with the purpose of evaluating the previous with the present performance. In this sense, for Brown, the concepts of assessment and evaluation have similar meanings. For Bachman and Palmer (2010) and Hyland (2003) assessment and evaluation hold different conceptualizations. They describe assessment as the process of collecting information about a specific area, which results in a score or a verbal description that may possibly be used to `evaluate´ students (Bachman and Palmer, 2010, p.21; Hyland, 2003). In other words, students are evaluated when important decisions are made based on the assessments carried out in the classroom. When assessing, language teachers are interested in making judgments of students´ or test takers´ language proficiency while seeking to fulfil specific assessment purposes such as a) screening and identifying, b) placing, c) reclassifying, d)

monitoring student progress, e) evaluating programs, and f) accounting for the program. (O´Malley and Pierce, 1996). For the purposes of this project, the concept of assessment given by Bachman and Palmer (2010) and Hyland (2003) will be taken into consideration.

According to Scarino (2013) and Lam (2015), in terms of assessment and what it involves, two paradigms can be considered. The first, the traditional view, considers assessment a cognitive aspect of learning and psychometric testing where teachers focus on assessing students´ learning (assessment of learning). The second, focuses on a sociocultural view of learning where the importance of assessment lies in the contextualization and the social interaction of those being assessed and those assessing (Ibid, p.312). Usually, Mexican institutions promote the traditional view of assessment encouraging summative assessment of learning therefore representing a sense of uneasiness among language teachers. On the one side, instructors understand the knowledge and importance of authentic assessment but are required by their institutions to comply with traditional views of assessment. For language instructors to give assessment the importance it represents, it is necessary for them to find ways of combining a traditional view of assessment, to comply with their institution´s requirements, and a more alternative view of assessment to favour language development. However, this may represent tension among teachers, especially if they do not have the adequate training. The importance of assessment may lay in the implications and consequences it has in the classroom for language teachers, language students and maybe even for human lives. As Deborah Crusan (2010, p.p 8-9) states,

> Assessment is everywhere. We perform assessments all the time. We make assessments (or judgments) about hundreds of things, big and small…assessment assists us in making all kinds of decisions, helps us grow intellectually and socially, and maybe saves our lives.

For the EFL writing student and teacher, assessment plays a crucial role for language development. For students, writing is one of the main language skills that allows them to engage in several stages (brainstorming, drafting, revising, editing, among others) in which they constantly need to be assessing their text for further improvement (Ibid). In other words, assessment allows students to reflect on their work and implement suitable remedies for meaning to be convened. For the language instructor, assessment is a duty (Weigle, 2007) that needs to be taken seriously. After all, every education institution or language program requires their teachers to assess their students in a specific way. Teachers need to have knowledge on how to create, implement, administer, assess and finally communicate results to test takers as valid and reliable as possible (Weigle, 2007) therefore giving assessment the importance it has for students.

Linda Taylor (2009) adds that not only teachers need to have this literacy to test and assess language skills, but also national examination boards, academics and students engaged in assessment research, language teachers, or instructors, advisors, decision makers within language planning and education policies, parents, politicians, and general public involved in language assessment (Ibid, p.25). In other words, assessment literacy is of major concern to every type of assessment and testing stakeholder.

The assessment of writing may have additional consequences for students´ and teachers´ lives. Washback, the effects that assessment can have on teaching and learning (Brown, 2007; Hyland, 2004), may not only have learning and teaching consequences but can also result in academic changes that can bring upon positive or negative effects. In the Mexican context (and many other contexts in the world) these changes may imply students obtaining or not a degree diploma, or the successful or unsuccessful enrolment in a

language level or university program. Ideally, every decision and its corresponding consequence should be beneficial for the student, the teaching and the learning environment therefore resulting in a nonlinear connection among assessment (collection of information), evaluation (decision making) and the teaching and learning (resulting consequences). As shown in Figure 4 titled "Assessment, Evaluation and Teaching & Learning" without having clear assessment objectives, decisions cannot be made and consequences may negatively impact students´ language development (Bachman and Palmer, 2010, p. 27).

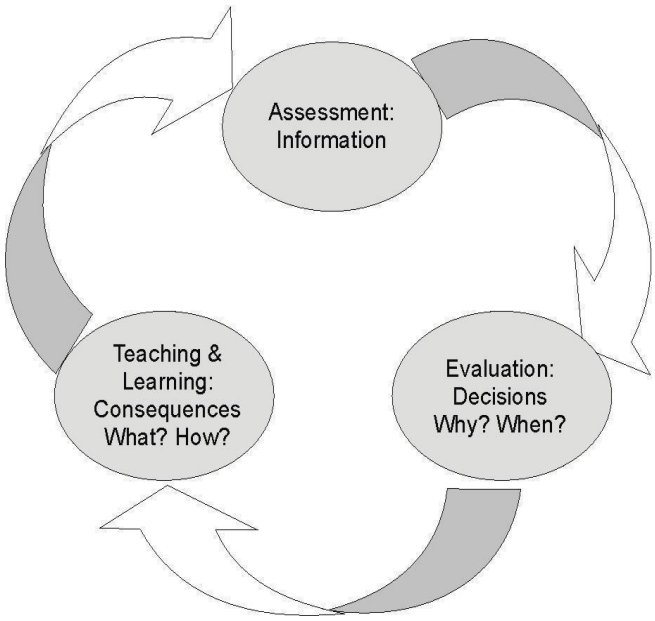

Figure 4 Assessment, Evaluation, and Teaching & Learning (Bachman and Palmer, 2010, p. 27)

Finally, it is important to consider that assessment may provide language teachers a window to understand their students and their own teaching practice. As Huot (2002, p.20) states, `my message to teachers is that the proper and intelligent use of assessment can provide them with the opportunity to learn rich, useful information, about their students, pedagogy, and programs´. Therefore, allowing improvement to be a result of language

assessment. The following section provides a description of some crucial characteristics of language assessment.

2.4.1 Characteristics of Language Assessment

Assessment is systematic and of substantive grounding (Bachman and Palmer, 2010). In other words, it follows a specific process and is based on the specific content of the language program. By collecting information about students´ learning or the language program, a strong bond among individuals, the learning environment and potential consequences is built. As portrayed in Figure 5, once the test taker or the language student is engaged in an assessment task, a score or description is given. This interpretation of the assessment record leads to an interpretation of the students´ language ability, which results in a decision-making process. Finally, this decision results in life-changing consequences for the assessed student. Certification of professional employment, a placement in a specific language level of a specific language course, and the need to improve an existing program are examples of important decisions made based on the assessment data (Ibid).

When assessing writing, context is an important factor to consider. Edward White (1990) describes writing assessment as a field that largely depends on the environment that surrounds the writer. In other words, when assessing writing, the specific students, the specific program, the specific learning and teaching goals need to be considered in assessment practice. Huot (2002) considers that writing assessment should be `site-based and locally controlled´ (p.19). For this to happen, teachers need to consider assessment part of their everyday practice and adapt it to their everyday needs and situations. Crusan (2010, p.12) comments in this sense that one type of assessment does not suit every student, every teacher, and every institution.

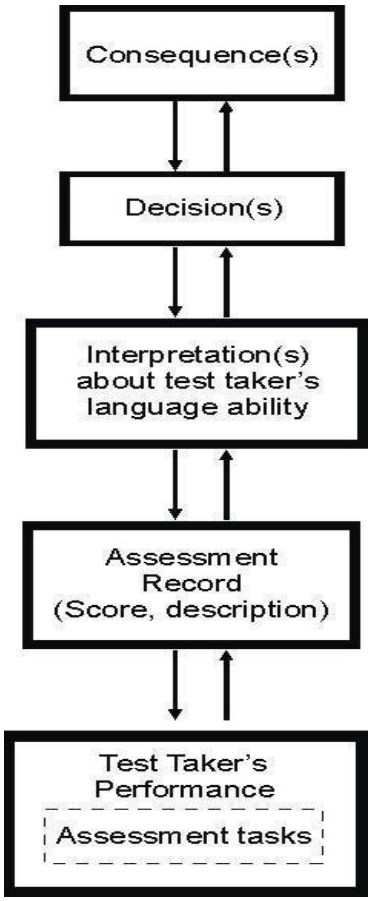

Figure 5 Test Taker´s performances to intended uses: Decisions and Consequences (Bachman and Palmer, 2010, p.23)

Therefore, assessment strategies and principles followed in the classroom depend on context specific issues and will benefit if specific principles are considered when developing assessment standards. The section below points out important assessment principles to consider when assessing a language skill, specifically reliability.

2.4.2 Reliability as an Assessment Principle

Whether assessment takes place in the classroom or in large-scale testing, specific principles should be followed. Brown (2007) considers that five principles should be considered for language assessment, for instance a) practicality, b) reliability, c) validity, d) authenticity and e) washback. Crusan (2010) adds to these principles that assessment

should a) be transparent to students, b) follow previously stipulated standards, and c) seek to fulfil specific goals and objectives. Ken Hyland (2003) comments that principles of assessment such as validity and reliability should be attended and cared for during writing assessment.

The field of writing assessment is a difficult one to explain and understand when it is considered that a single writer may produce inconsistent pieces of writing and raters may assess a single piece of writing in different ways. As stated by Pearson (2004, p.124),

> …it is widely believed that the only fair way for writing to be assessed is if all readers agree on the scoring criteria and then respond in ways that are similar to each other and that are consistent within their own readings. If raters A and B disagree on how to rate an essay, how can the final score (e.g., an average score or a total) be fair or meaningful to the writer? Similarly, if a rater scores one way when fresh and another when fatigued, how can a student whose paper is read when the reader is tired be rated fairly?

Barbara Kroll (1998) considers that reliability of writing assessment depends on the extent to which various raters give the same score to a single text. Therefore, giving allusion to the importance of inter-rater assessment and to the differences that can be found among distinct assessors. Therefore, rater variability and reliability are part of the issues of writing assessment that researchers and above all language teachers face in their everyday practice.

Reliability of scores refers to the consistency of scores: inter-rater and intra-rater consistency. In other words, the ability of a test score to be replicable from one test occasion to another (Hamp-Lyons, 2003). While inter-rater reliability refers to the consistency with which different scorers assess the same paper, intra-rater reliability refers to the stability with which a single rater scores the same paper on different occasions (Bachman and Palmer, 2010; Brown, 2007; Fulcher and Davidson, 2007; Weigle, 2002). It

has been stated by experts that 100% of writing assessment reliability is impossible to obtain, however 75% is now obtainable in distinct parts of the world (Hamp-Lyons, 2003). Fulcher and Davidson, (2007, p.131) add that,

> It has been acknowledged since the beginning of what used to be called `subjectively score´ tests that it should be a matter of indifference to a test taker who scores the performance. If the score is likely to change depending upon the rater, the question arises as to whether it is the rater´s own personal views that impact on the score, rather than the ability of the test taker.

Reliability of scores can be interpreted with a series of statistical analysis calculations such as average score (Mean), Standard Deviation, Correlation Coefficient or ANOVA analysis (Weigle, 2002).

For the purposes of this study, reliability is considered a factor that reflects the consistency with which teacher participants provide a holistic and an analytic score to written samples (Kroll, 1998) on two separate occasions (Bachman and Palmer, 2010; Brown, 2007; Fulcher and Davidson, 2007; Weigle, 2002). This consistency may also be reflected in the teachers´ use of rubrics and their interpretation of scale descriptors considering their academic preparation, their teaching experience and other contextual traits that may have a role in teachers´ inter-rater reliability. Although, this project is not developed under a large-scale testing context (where reliability calculations are more frequently run) it is considered that by comprehending the link between score reliability levels at a classroom level, the training sessions conducted and its influence in teachers´ interpretation of the holistic and analytic tools used, the impact of training on classroom writing assessment can be further understood. This study analyses the inter- rater reliability prior and post to training considering participants´ gender, teaching experience, and academic preparation

by running statistical calculations such as Reliability Analysis, an Independent t-test and a Paired Sample t-test.

Classroom contexts involve a variety of different scoring procedures and assessment tools such as scoring rubrics that teachers can use depending on their assessment purposes or their students´ needs. However, the assessment of writing in large-scale testing and classroom contexts has not always been the same. As time has evolved, assessment has changed as well. The following section provides a description of the historical development of writing assessment.

2.5 Historical development of Writing Assessment

China began with formal testing in the Sui Dynasty with the Imperial Examination (581-619) which had the purpose of selecting people for high ranking bureaucracy positions without considering their social status (O´Sullivan, 2012). Later on, in the Chou (1111-771) period essay testing (Direct testing) began (Hamp-Lyons, 2001) and direct actions to care for assessment reliability were implemented. Locking up candidates and examiners in the same room were actions with which education authorities expected to assure standardized and reliable assessments (Hamp-Lyons, 2001). Europe also implemented direct essay testing in addition to their already implemented oral examinations. Once Britain became aware of the need of literate officials to administer their colonies among the world, universities began providing education to more and more people thus assessing large amounts of written texts. However, having more written texts to assess raised the question of standardization, validity and reliability among assessed texts (Grabe and Kaplan, 1996; Hamp-Lyons, 1990; Huot, 2002).

Meanwhile in the United States, Harvard University (1873-1874) was the first to include a written examination in their entrance requirements (Hamp-Lyons, 2001). Their increased preoccupation for standardized assessment led to the implementation of multiple-choice (Huot, 2002) examinations that could reflect the true ability of a student (Hamp-Lyons, 2001). Multiple-choice questions were created by Fredrick J. Kelly in 1914 (O´Sullivan, 2012; Fulcher, 2014) in an attempt to give solution to issues arising. In a time when education was rapidly growing in the United States, teacher scoring subjectivity and teachers´ lack of time to mark a growing number of papers were two of the main issues that needed to be diminished. Multiple-choice questions allowed assessment to be cheaper and become more standardized (O´Sullivan, 2012). Attempts to standardize the assessment of handwriting began in 1908 with Edward L. Thorndicke when he created the first standardized test (O´Sullivan, 2012). Milo B. Hillegas continued with Thorndike's assessment methodology in 1912 and created the first scale to assess English Composition. Two years later S.A Courtis (1914) implemented the first English standardized test (Ibid).

The birth of multiple-choice questions and standardization opened the doors for the creation of many language tests such as the Test of English as a Foreign Language (TOEFL) created by the Educational Testing Service (ETS) in 1947 which implemented multiple-choice items (Grabe and Kaplan, 1996; Hamp-Lyons, 1990, 2001). Other tests such as the Test of English for International Communication (TOEIC) by the ETS, the International English Language Testing Service (IELTS) created by Cambridge English Assessment and the British Council´s recent created Aptis test also follow multiple-choice formats and standardization processes.

In 1986, Fader (1986 cited in Hamp-Lyons, 2001) and other educationalists argued against this type of examination exposing direct testing as the only way of visualizing true student development within this skill. Classroom teachers argued that indirect testing of writing led to a college-entry society that could not think critically (Hamp-Lyons, 1990, 2001) and by the 1970s universities reintroduced written examinations into their entrance requirements. The University of Michigan was the first to introduce written tasks as entry requirements and nowadays it holistically scores written examinations (Grabe and Kaplan, 1996).

ETS also felt the pressure and in 1986 introduced the Test of Written English (TWE) as an option to take with their TOEFL examination (Hamp-Lyons, 1990; Kroll, 1991). Although direct testing of writing has been favoured in opposition to indirect writing assessment throughout time, the latter did set an important benchmark with which current writing assessment can account for: reliability and validity measures (Grabe and Kaplan, 1996). When designed carefully, indirect writing tests found higher correlations in terms of reliability standards in comparison to direct writing assessment, which nowadays is more concerned with construct and content validity. As stated by Grabe and Kaplan (1996, p.399),

> The increasing emphasis on construct and content validity-whether the test reflects what research understands writing to be, and what is normally covered by writing practices- will push future writing assessment further towards direct assessment approaches

Although for large-scale language test developers, standardization issues were solved with the implementation of indirect writing tests, experts have come to suggest that these tests did not reflect student language users and test takers true ability to interact with writing skills, such as the awareness of audience, the implementation of coherence and organizational patterns, among others.

In the last decades, a considerable amount of attention has been provided to the assessment of writing. In the 1990s, the creation of journals that entirely devoted their attention to writing assessment contributed to the expansion of the field. Journals such as Assessing Writing and The Journal of Writing Assessment (Huot, 2002) as well as the Journal of Second Language Writing approach issues and trending topics in the field of writing assessment. However, other journals also approach issues in language assessment such as Language Assessment Quarterly, Language Testing, Assessment in Education: Principles, Policy and Practice, Papers in Language Testing and Assessment, among others. These journals focus on the always evolving field of language assessment and the assessment of writing.

Nowadays, current assessment and testing trends are constantly being questioned by experts in the field. For instance, researchers have pointed out the importance of considering not only the collection of test takers´ scores but also considering contextual traits of assessment such as material developers´ views, stakeholders´ assessment literacy, stakeholders´ experiences in assessment, the local culture of learners and assessors among others (Inbar-Lourie, 2017; Yan, Fan and Zhang, 2017; Wang and Yan, 2017). Another shift which is worth bringing forward is the focus that language testing and assessment research have now approached. Most of the research has focused on large-scale tests, their design, implementation or scoring. However, research is turning to other important aspects of language assessment such as teachers´ conceptualization of assessment and the need for assessment literate teachers. In other words, it is necessary to approach assessment from a multidisciplinary stance to understand and improve assessment (Inbar- Lourie, 2017). Finally, technology and its involvement in the assessment process in contexts where

massive numbers of test takers (such as China) need to be assessed have also been a turning point in research (Jin *et. al*, 2017). Technology such as computer-delivered tests, machine-scoring and online teacher training are factors that have recently emerged as variables with an active role in language assessment. The following section further describes such variables, the current assessment trends and the status of testing and assessment in Mexico.

2.6 Current Writing Assessment Trends

Nowadays, some experts believe that a new era, the fourth generation to writing assessment needs to be introduced. Hamps-Lyons (2001) considers that this new era of writing assessment needs to be recognized for its qualities such as a) the use of technology, b) a humanistic view of `great fairness´, c) a political view in which test developers, curriculum developers and government stakeholders take responsibility for the test they develop and finally d) an ethical view in which it is important to consider what is fair for the test taker and what is the fairest for the mass that takes the test. Language teachers need to keep in mind the characteristics of the new era of writing assessment so that these can enlighten their everyday assessment practice in benefit of their students. However, in Mexican EFL classrooms the assessment of writing is a difficult task to carry out for many reasons. Issues such as lack of teacher training and an overload of curriculum content, limit the amount of time teachers have in the classroom to develop appropriate language assessment. Therefore, making fourth generation assessment difficult to attain.

In Mexico, language proficiency tests that use direct writing tasks are required for different reasons. In tertiary levels of education, undergraduate students are required to prove a specific level of proficiency of a foreign language, being English the most popular among

the languages chosen. In Tamaulipas, a state situated in the north-eastern region of Mexico, the Universidad Autónoma de Tamaulipas (UAT, Spanish acronym for Autonomous University of Tamaulipas) from 2007 and on requires all of their undergraduate students to prove a B1 CEFR level of English proficiency for them to obtain their undergraduate diploma (Universidad Autónoma de Tamaulipas, 2011). The Centre for Languages and Applied Linguistics (situated in Ciudad Victoria and Reynosa) and the Centre for Foreign Languages (Tampico) of the University provide students with distinct options to fulfil this requirement. The Test of English of Tamaulipas (EXIT, Spanish acronym for Examen de Inglés de Tamaulipas) is a test created by the Centre that has the purpose of assessing students' English proficiency for graduation purposes. Another locally made examination accepted is the Exam of the University of Veracruz (EXAVer, Spanish acronym), which was created by the University of Veracruz with the support of the British Council and the University of Cambridge and assessment experts such as Professor Barry O´Sullivan (Dunne, 2007). Another option provided is the TOEFL exam in its Institutional (ITP) and Internet (IBT) versions, thirdly the Test of English for International Communication (TOEIC) and the First Certificate of English (FCE).

In the EFL classroom, teachers are required to assess the four language skills simultaneously with the use of different methods. At some institutions, teachers are required to hand in scores of language skill development on a monthly or bimonthly basis that are mostly obtained from written tests. Most of the institutions leave the creation or adaption of such tests to the teacher. They design, adapt, interpret and report student scores to the institutional office. Other institutions provide the assessment tool to the teacher (exam, task, prompt, among others) so that it is conducted, scored and interpreted. These instruments, either created by the classroom teacher or the language program

administration, mostly include reading, listening, and grammar and vocabulary components. Speaking and writing, because of its productive nature and its complex assessment process, on many occasions are avoided or given a less percentage of the final score.

Although experts consider that language assessment needs to be context specific (Scarino, 2013; Weigle, 2002; White, 1990), and that the assessment purpose of each institution and program are of great importance, distinct institutions were considered for this study. The purpose of having distinct research contexts was to have a scope of the different writing assessment practices and to compare how participant instructors differ in their reported assessment practice. Additionally, teacher participants of this study were recruited from different institutions to fulfil the necessary sample to interpret the data obtained. These participants and institutions are further described in Chapter Four of this thesis. To be able to understand the assessment context of EFL teachers, it is necessary to understand the potential issues faced when assessing writing.

2.7 Assessing Writing in the ESL/EFL Classroom

According to Grabe and Kaplan (1996), writing assessment generally occurs in two different contexts: classroom environments and large-scale standardized testing contexts. While classroom assessment can be used for diagnostic, placement and achievement purposes, standardized testing focuses on proficiency judgment purposes. One of the main differences among classroom assessment and large-scale testing is the direct contact with the student (Fulcher and Davidson, 2007). The classroom is considered a social situation, which is largely based on how people interact in this environment and others in it (Ibid, 2007). A large-scale testing context does not leave room for this consideration. Instead, test

takers interact with a specific room in which a test is answered and in which aspects such as temperature, colour, decoration, invigilator (test administrator) are involved in the test taker's test results. If context factors are involved in test results, then it is considered that the validity of scores is construct-irrelevant (Ibid).

In the language classroom, the context is not irrelevant to the development of students. `How well they (students) are progressing can be assessed only in relation to their involvement with the context and others with whom they interact in the process of learning. The context is part of the construct´ (Fulcher and Davidson, 2007, p.25). Weigle (2002) outlines a set of differences among large-scale testing and classroom assessment. Those that stand out are a) for large scale testing a numerical score is all that is needed while in classroom assessment the numerical score is frequently accompanied by feedback; b) people taking the examination are brought together on the same date and same time period while in the classroom time is more flexible; and finally, c) large-scale testing focuses on language users of many different backgrounds while in the classroom the instructor must consider the specific learning context.

In the language classroom, assessment can be carried out in two modes: implicit or explicit. While during implicit assessment the process may be merged with teaching, in explicit learning an important distinction is made between assessment and teaching (Bachman and Palmer, 2010). During implicit learning, formative decisions (to correct or not students´ response) that may lead to student improvement are taken, in explicit learning summative decisions are made without the intention of forming or helping students improve. Another important difference is that while being assessed implicitly students and

teachers may not be aware of the process; in explicit assessment, the participants are consciously aware that assessment is taking place.

This leads to the difference among summative and formative assessment as pointed out by Ken Hyland (2003). While the former focuses on improving students´ performance and identifying their strengths and weaknesses with the purposes of improving any flaws, the latter focuses on `summing up´ (Hyland, 2003, p.213) the amount of knowledge the student has learned in a period of time. In other words, formative assessment may take place during the teaching practice in the classroom while summative takes place at a specific period of time to set forward individual student accomplishments or outcomes (Huot, 2002).

Regarding the assessment of writing, it can be done directly and indirectly (Grabe and Kaplan, 1996; Harmer, 2007). Direct and indirect writing are the types of tasks used both in classroom and standardized testing contexts. Tasks in which test takers are required to write a sample of a text are considered direct assessment. In other words, performance assessment requires students to perform or produce any form of task orally or in written form that allows them to recall previous knowledge, recently learned information or relevant skills (O´Malley and Pierce, 1996). Tasks such as multiple choice or cloze activities are examples of indirect assessment (Grabe and Kaplan, 1996; Hamps-Lyon, 2003; Harmer, 2007).

2.7.1 Formative Assessment in the EFL Classroom

Formative assessment, according to Brown and Abeywickrama (2010) refers to the process of `forming´ (p. 7) students in the classroom with the purpose of improving their language

use. In other words, classroom assessment that follows a formative purpose intends to assess language abilities and decide what to do next as a teacher and how further learning can take place (Fulcher and Davidson, 2007; Yorke, 2003). It can take place in a formal setting, such as those where students are required to submit specific tasks as part of the curriculum being taught and are provided by their teachers with suggestions to improve; or in informal ways (casually analysing the content that has been covered in a day´s lesson and deciding what needs to be covered once again) where the course of events guides the assessment (Yorke, 2003).

To contrast, summative assessment has the intention of measuring what a student has managed to learn at the end of a term, course, unit, etcetera (Brown and Abeywickrama, 2010). It seeks to portray if learners have accomplished goals but does not necessarily raise awareness of what needs to be improved. It usually occurs in a formal setting because the school curriculum requires scores to be handed in without seeking for student improvement. Examples of summative assessment are proficiency exams or final course examinations that may involve decision-making or not (Ibid).

However, there has been discussion among researchers of the difficulty of developing formative assessment in the classroom and the effect that large-scale summative assessment and their corresponding national policies may bring upon the assessment practice of classroom teachers. In a first language (L1) context, Black and William (1998a), for instance, consider classroom assessment in the American educational context, as a black box in which different types of variables are added such as tests, assessment tools, contextual factors and teacher/student social factors. Then, specific output is expected. Usually, it is expected for students to obtain satisfactory results in high-stakes

tests. However, little attention is paid to the actual process of learning and teaching that happens inside the box that contributes to the output generated. As a result of their documentary analysis of 580 journal articles or book chapters, Black and William (1998b) concluded that classroom assessment in the United States encountered difficulties such as the negative impact that assessment had on students´ learning, the overemphasis of quantity of students´ work over their quality in feedback and the collection of scores or marks seemed to be prioritized over students´ development. The researchers pointed out that these constraints classroom assessment is facing are a product of national assessment standard policies implemented in the US education system in the early 90s. Nevertheless, it is pointed out that the best way to improve classroom assessment is by improving the conceptualization that students have of assessment, since they are the primary users of formative assessment developed in the classroom; and that the implementation of classroom assessment `…calls for rather deep changes both in teachers´ perceptions of their own role in relation to their students and in their classroom practice´ (Black and William, 1998b, p. 20).

Another issue with assessment that has been widely discussed among researchers is the difficulty of formative and summative assessment in higher education. For instance, Yorke (2003, p.480) points out the fuzziness of the limitations of these concepts and states,

> Some assessments (e.g. in course assignments) are deliberately designed to be simultaneously formative and summative –formative because the student is expected to learn from whatever feedback is provided, and summative because the grade awarded contributes to the overall grade at the end of the study unit. Summative assessments in relation to a curricular component (the student passes or fails a module, for example) can act formatively if the student learns from them.

Rea-Dickens and Gardner (2000) as well as Lee (2007) discussed this difficult conceptualization of formative and summative assessment. The former stated that the difference did not only lay on the intentions of the teacher and the amount of feedback provided but it also lays in the importance of assessment decisions that language teachers make on a daily basis and the implications that these may bring to high-stake test results. They conclude in their study, that although research has considered classroom assessment as low stakes in comparison with language testing, the importance of the decisions made by teachers on a daily basis are of `high-stake´ importance. It is teachers who decide which students are promoted to continue their language studies or who are not. In the Mexican EFL context the situation described by these researchers is very similar: teachers in the classroom decide who moves on and who does not, they are unaware of the adequate procedures to collect language samples or of how to make valid and reliable decisions of students work (Rea-Dickens and Gardner, 2000). Finally, most of the Mexican EFL teachers working in the tertiary context do not have assessment training and instead assess students´ language by instinct or day-to-day practice (Metler, 2003; Koh *et al.,* 2017).

The difficulty that formative assessment represents in the classroom becomes more complex when carried out in an ESL or EFL classroom. Focusing on fourth grade elementary level English as an Additional Language (EAL) students in London and two in-service teachers, Leung and Mohan (2004) analysed the social aspect of classroom assessment and found that formative assessment allows teachers to consider the decision-making processes that students experience while their speaking skills are being assessed. The authors suggested that formative assessment should emphasize `…student processes as well as products, … student-student interaction, …teacher use of scaffolding and, in particular… student decision-making discourse, all under locally adaptive conditions´

(p.343). They bring forward the importance of the teacher in the `assessment for learning´ (Leung and Mohan 2004, p.337; Klenowski, 2009, p.p 1-2) process and consider that classroom teachers have a great challenge to face while being expected to perform formative and summative assessment in their classrooms (Vogt and Tsagari, 2014). It is clearly stated by the authors that formative language assessment raises more questions and issues in comparison with standardized assessment. As in the British context, Mexican EFL teachers are expected to perform a series of assessment processes during which they are required to collect evidence, judge evidence, score evidence, interpret, report to administrative offices every specific period of time and make decisions to improve based on the assessment collected therefore combining both types of assessment: formative and summative (Vogt and Tsagari, 2014). Very frequently teachers are not academically prepared to collect evidence from students in order to make relevant decisions. Therefore, research still needs to inquire on the usefulness of assessment training and how it can improve teachers´ regular assessment practices in the EFL classroom.

Lee (2007) builds on the concept of `assessment for learning´ by considering it a process that seeks to use student evidence to be interpreted and then used by learners and teachers to make decisions about learning. For instance, where teachers and students currently are, where they need to go and what strategies need to be implemented to get there (Assessment Reform Group, 2002; Lee, 2007). Specifically, for ESL/EFL writing, Lee (2007, p.p 202-208) considers that some important principles should be cared for when seeking to improve learning and teaching through assessment. For instance, some of these include, a) sharing learning goals with students, b) helping students understand the standards they are working towards, c) involving students in assessment, d) providing feedback to improve the text, e) assimilating mistakes as a natural part of classroom

engagement and learning, and f) integrating teaching, learning, and assessment in writing. However, it is specified by the author that one of the issues faced by language teachers when assessing writing with this purpose is the institutional requirement of scoring papers, which may lead to student demotivation. Finally, the author concludes that teachers need to have support from their institutions to understand the use and applications of assessment in their classrooms by stating,

> …teachers work collaboratively to review their writing instruction practices and plan a comprehensive program that takes into account the interrelationships between teaching, learning, and assessment. They can then develop strategies to teach writing and formulate a clear feedback policy in the light of their writing program…. (assessment for learning) should be considered a key professional skill for teachers, who need support through continuing professional development. There are significant implications for teacher education in helping teachers come to grips with AfL (assessment for learning) in writing (Lee, 2007, p. 209).

In an attempt to define EFL classroom assessment, Ketabi and Ketabi (2014) considered that the difficulty lies in the nature of assessment that teachers perform in their classrooms. They may at times assess explicitly and implicitly, or formatively and summatively during the same course depending on their teaching needs, their students' needs and the administrative needs of their institutions. There are contexts, like the EFL Mexican context, in which teachers are required to submit summative test scores to the institution administration at the end of a specific period of time without providing any type of formative feedback while on other occasions feedback may be extensive and scores are not necessary (Ketabi and Ketabi, 2014). Additionally, the problem is more notorious when in crowded classes the teacher does not have enough time to provide insightful feedback to the student therefore collecting information through tests.

Little is known about what teachers actually do in the ESL/ EFL classroom in terms of their assessment practices (Polio and Williams, 2011). However, researchers such as Davidson and Cummings (2007) consider that much of the assessment carried out in the classroom is designed and implemented by the classroom teachers thus involving them in the planning, developing and judging of tasks and students´ performance. Students may also be involved in this process, especially when engaged in peer or self-assessment (Ibid). Other experts such as Cumming (2001), Cheng *et al*. (2004), Cheng and Wang, (2007), Cheng *et al*., (2008) and Inbar-Lourie and Donitsa-Schmidt (2009) focus on describing how ESL/EFL teacher classroom assessment is carried out.

Cumming´s study (2001) interviewed 48 experienced EFL professors and found that depending on the purpose of their courses, whether a specific or general purpose English writing course, their rationale for choosing a specific assessment task and setting specific assessment standards changed. Those that taught specific English courses used limited amount of assessment criteria as well as limited forms of assessment. On the other hand, those teaching general purpose English writing had more varied ways of assessing students as well as more varied standards of assessment.

Another study developed by Geoff Brindley (2001) approached the problematic relationship between outcome assessment and the evaluation reporting systems of Adult immigrant education systems in Australia. He approached the issues of validity and reliability in teacher constructed assessment tasks and the assessment standards set for these assessments. He stated that in the case of English writing, set tasks were varied and that teachers did not agree on standards to consider for evaluation (Polio and Williams, 2011). The study concludes by suggesting the use of a bank of piloted and benchmark tasks

so that teachers can have tasks to choose from and avoid losing time in their creation. He also considered that professionalization and teacher development are crucial for the success of outcome assessment based language programs. Therefore, the importance of analysing how assessment training may or may not lead teachers to improve their regular classroom writing assessment practice is emphasized by this author.

Other experts have attempted to understand classroom assessment in foreign language (FL) contexts by comparing the methods used in different parts of the world. While Cheng *et al*., (2004) focus on teachers from Canadian and Chinese contexts; Inbar-Lourie and Donitsa-Schmidt (2009) focus on Israeli teachers and their assessment procedures. With the purpose of conducting a comparative study among the assessment practices of teachers from five different parts of the world (Alberta, British Columbia, Ontario, Hong Kong and Beijing), Cheng *et al*., (2004) collected the answers that 95 (51.3%) Canadian teachers, 45 (32.0%) teachers from Hong Kong, and 124 (95.3%) from Beijing gave to an online survey that explored the assessment methods used by teachers in several language aspects. Quantitative data analysis indicated that Canadian instructors identified the most purposes for assessing writing than their Hong Kong and Beijing counterparts, being the short essay the activity most widely used (85% of participants) to assess EFL writing. Other assessment tools used were student journals, and portfolios which were mostly used in Canada. Hong Kong and Beijing reported to use the long essay in addition to sentence and paragraph editing as teacher created activities while portfolios were among the student-centred activities.

In a more qualitative-oriented comparative study and focusing on six different areas of assessment, Cheng et al., (2008) interviewed 74 teachers from Canada, China and Hong

Kong who had previously participated in their 2004 study (Cheng et al., 2004). The interview explored the areas of (1) Developing and Choosing Methods for Classroom Assessment, (2) Judging and Scoring Student Performance, (3) Reporting Final Course Grades, (4) Impact of External Testing, (5) Education and Training in Classroom Assessment, and (6) Background Information (Ibid, p.11). The three contexts had in common the source of their assessment methods; assessment activities were created by the teacher or obtained from printed materials, which provided standardized test preparation to students. Specifically, writing and oral skills were tested with the use of these materials; Chinese professors were the ones who mostly used them at the end of a school period. Although Canadian teachers used these methods as well, they mostly used summaries and short essays throughout a whole course. China was characterized by the use of translation as an assessment activity created by the teachers to prepare students for the standardized English tests that they would need to take in the future.

In a very similar comparative study, Inbar-Lourie and Donitsa-Schmidt (2009) analysed the assessment perceptions and alternative assessment classroom procedures of 113 EFL teachers with the use of a self- report questionnaire. The researchers considered that contextual and institutional factors had a great impact on teachers´ assessment procedures. Therefore, taking into consideration that Hargreaves, Earl and Schmidt´s (2002 cited in Inbar- Lourie and Donitsa-Schmidt, 2009) model may help portray teachers´ perceptions and beliefs. The researchers analysed participants´ views within four major perspectives – technological, cultural, political and postmodern (Ibid, p.188). Regarding the technological aspect, time management, organizational factors and resource availability were considered. Cultural traits were those that reflected the integration of assessment procedures within the specific sociocultural environment of the school. It also accounted for the collaboration

among different stakeholders during the assessment process. A third element were political factors that may have a strong washback effect in teachers´ classroom assessment (Froetscher, 2017): top-down monitoring, standardized tests, bureaucratic interference or institutional policies. Finally, post-modern traits were attributed to how teachers viewed the `uncertainty that characterizes the present era, thus critically questioning the authenticity, reliability and validity of assessment beliefs and practices´ (Inbar- Lourie and Donitsa-Schmidt, 2009, p.188). It also included the role of authentic assessment as an innovating method to assess language. The researchers´ concluded that the teacher participants highly agreed with the use of a variety of alternative assessment techniques while a high level of agreement was also found for the technological obstacles that alternative assessment faced. The Technology and Post-Modern perspectives were considered the areas that hindered the teachers' implementation of alternative assessment techniques (teachers considered students cheated when submitting work). Finally, it was concluded that the amount of time that teachers need to invest in alternative assessment and the lack of training were two important factors that teacher participants of this study find as determining to avoid using alternative assessment. This finding is very similar to that encountered in the Mexican EFL context. In the English classroom, teachers very frequently avoid assessing writing. They do not feel professionally prepared to conduct assessment and the heavy workloads that teachers face regularly diminish the amount of time teachers can dedicate to writing assessment. Although these similarities were found among the Mexican context, it is my belief that a qualitative component to the research methodology of this project could be included to further understand teachers´ views and the specific context that may have a role in these views.

Chen *et al*., (2013) focuses on the analysis of the sociocultural conditions of the Chinese context (economic, social and cultural factors) and how these impact the assessment of English in the university classroom. The perceptions of the participating English teachers allowed portraying the problematic situations that teachers face when assessing language in their classrooms. After interviewing a language program administrator, two classroom teachers and conducting one teacher focus group at each of the two public universities participating in the study, data was analysed and transcribed. Analysis indicated that the significant differences found among the perceptions and practices of the teachers mirrored the `institutional culture´ (p.15). It allowed comprehending that teachers' experience, training, student expectations and conceptualization/operationalization of assessment were also products of the institutional culture that is present in the university. Although this study provides a unique perspective in regard to teachers´ views of assessment in their Chinese context and the difficulties encountered, it may be enriched if students´ perspectives were also considered to complement the views of involved stakeholders. By understanding students´ voice, teachers may also consider the learning environment in which assessment is immersed.

In the Mexican context, such as the context described by Ketabi and Ketabi (2014) and Inbar-Lourie and Donitsa-Schmidt (2009), teachers are required to perform formative assessment of students' language skills as well as a summative assessment every specific time period. These `teacher´ activities are bound to the institutional policies of each school and which instructors are obligated to comply with. However, the assessment of writing has followed two distinct cultural practices: some university institutions incorporate writing in their periodical assessment while others omit its assessment. Some institutions use scoring rubrics to provide a summative score as well as to provide feedback to the

students in relation to how the text can be improved. In other words, EFL teachers that assess writing use analytic and holistic scoring rubrics as tools to assess classroom writing (Brown and Abeywickrama, 2010). Institutions such as one of the participating language institutes of this study implement rubrics that have been previously adapted from other existing rubrics which were created by assessment experts such as Weir (1990) or the British Council. Other teachers from different university contexts, use rubrics that are available online and comply with the criteria included in them without any adaptation to their assessment context. Therefore, jeopardizing assessment validity. The following paragraphs describe the concept of rubrics and their role in the assessment of writing.

2.7.2 Scoring Rubrics: Tools for Classroom Assessment

Scoring procedures are of great importance, because `the score is ultimately what will be used in making decisions and inferences about writers´ (Weigle, 2002, p.108). Additionally, a rubric provides the instructor with standardized criteria to follow and may result in an increase in score reliability (Ibid, 2002). Another advantage is that the use of rubrics may allow instructors of the same proficiency levels and among the same language program to maintain consistency throughout their assessment. Finally, this tool may also allow the teacher to simplify the assessing activity by providing descriptions of criteria to which assign a number rather than to provide lengthy comments or correcting every grammar mistake (Weigle, 2002).

According to Brown and Abeywickrama (2010), rubrics are devices that are not a stand-alone alternative to assessment, but instead should be tools that can aid teachers to develop performance-based assessment effectively and responsibly. They consider rubrics to be beneficial not only for teachers but also for students because rubric oriented assessment allows writers to focus their efforts, produce higher quality work, obtain higher grades, and

lower anxiety levels (Brown and Abeywickrama, 2010, p. 128). However, the authors warn teachers that rubrics should be used with caution for they may be less exact in portraying the reality of performance because their simplicity may disguise the depth of development of students´ performance in the classroom (Ibid, 2010).

In the 1950s, the United States military used rubrics that were simple semantic differentiated statements (Fulcher and Davidson, 2007) that did not give specific information about the test takers proficiency. Scoring scales or rubrics were firstly created by the British Council thus contributing to rubric development by creating the first holistic rubric in the 1950s (O´Sullivan, 2012). These had the purpose of defining ` how good does the language have to be for this particular purpose and domain? ´ (Fulcher and Davidson, 2007, p. 96).

Scoring rubrics can be characterized by two distinctive traits, a) whether the rubric is exclusive to a specific task or if its generalizable to any task and b) if several scores are given to the text or a single score (Fulcher and Davidson, 2007). Based on these characteristics, the most common scoring approaches are a) holistic rating; b) primary-trait scoring and c) multiple-trait scoring such as analytic scoring (Grabe and Kaplan, 1996; Hamp-Lyons, 1990; Weigle, 2002).

Weigle (2002) considers rubrics serve different purposes and focus on different aspects of the written text. While primary trait scoring focuses on the specific most important traits of a specific written genre, holistic assessment focuses on the written text as a whole and focuses on several aspects of the text from an overall view of the rater. Scorers that use primary trait assessment focus on the specific context and characteristics of the text written

therefore creating a specific rubric each time a task needs to be assessed (Hyland, 2004). On the other hand, analytic assessment considered by other experts such as Hamp-Lyons (1990) as multiple trait scoring, focuses assessment on several aspects of the text rather than on a single holistic aspect.

In the Mexican EFL context, analytic and holistic scoring scales are used by teachers with two main purposes: 1) to provide feedback to the student about their written performance and 2) to provide a score when evaluating summatively written performance. This project intends to focus on the usage of analytic and holistic scoring scales for classroom assessment on behalf of EFL teachers; therefore, the following section provides a more detailed description of both of these scales as tools for classroom assessment of EFL writing.

2.7.3 Analytic Scoring Tools

According to Weigle (2002), analytic rubrics allow for more information to be given to the students, they include multiple scores therefore tending to be more reliable than holistic rubrics. The rater needs to read the paper several times with an analytic rubric and give L2 students feedback (Grabe and Kaplan, 1996) on specific aspects of writing such as content, organization, language use, etc. (Weigle, 2002).

Some important advantages of analytic scoring include that it is more useful to be used among inexperienced teachers and for L2 learners `who are more likely to show a marked or uneven profile across different aspects of writing…´ (Weigle, 2002, p.120; Bachman and Palmer, 2010). Another aspect that is important to consider is the fact that `it is more useful in rater training (to use analytic rubrics), as inexperienced raters can more easily

understand and apply the criteria in separate scales than in holistic scales´ (Weigle, 2002, p.115).

However, there exists the danger of validity issues. Research still needs to investigate the extent to which appropriate feedback is given on each scale or how genuinely the scale and score represent students´ writing abilities (Grabe and Kaplan, 1996). Additionally, it takes more time for a scorer to assess a paper with an analytic scale. Scholars in this area such as Hamp-Lyons (1990), Weigle (2002), and Weir (1990) put emphasis on the need to have clearly described criteria and an understandable description of the levels or sublevels of the analytic scheme to homogenize as much as possible assessors' interpretations. Most of the participating teachers in this study are not experienced in the assessment of writing and those that actually assess writing in their lessons use some type of analytic rubric. Therefore, this study uses an analytic rubric as one of the scoring tools participants use to rate the sample papers. Another rubric participants used to score papers was the holistic rubric. The following section provides a general background to holistic scoring.

2.7.4 Holistic Scoring Tools

Assigning a single score to a written text as a whole based on a first impression, rather than giving different scores to different categories, is the central premise to a holistic scoring approach (Bachman and Palmer, 2010; Hamp-Lyons, 1990; Weigle, 2002; White, 1990). Holistic assessment is mostly used in large-scale assessment however it may also be used in classroom contexts. In both contexts, a numerical score that ranges from 1-4 or 1-9 is given (Grabe and Kaplan, 1996; Hamp-Lyons, 2003) to a written text. Each score represents a level of proficiency and includes a short description of the corresponding

level. For reliability to be cared for when scoring holistically, experts consider that it is recommendable for scorers to hold group discussions and share their insights in terms of the scores given. Holistic scoring is a faster approach and therefore less expensive because the paper needs to be read only once and a single score is assigned (Weigle, 2002). It is believed that holistic scoring can benefit writers because it is designed to focus the reader's attention on the qualities of writing instead of its weaknesses. However, specific drawbacks can be identified such as the fact that a single score given to a text may not fully represent learners´ proficiency failing to represent aspects of writing such as syntax, vocabulary, organization, among others. (Bachman and Palmer, 2010; Weigle, 2002). Additionally, the description of each scale on a holistic rubric is more generalized than that found on an analytic rubric and may at times make it difficult to know what exactly a score actually reflects on the student paper (Bachman and Palmer, 2010). From my point of view, the essence of analytic scoring is to trigger teacher and student analysis by assessing a piece of written work with the guide of a specific number of detailed descriptors. This is a very different purpose from that followed by a holistic tool. Rather than describing to detail a specific descriptor, a holistic rubric seeks to generally describe performance to allow writing assessment to be more practical and feasible. Therefore, I consider that both rubrics have major differences that keep them from being merged to construct a single rubric. Nevertheless, I believe they may be combined by teachers in their classroom assessment activities at different points of the year or semester.

Once the type of scale that suits the teaching and assessments needs of the teacher and the student is identified, it is necessary to create or adapt the scoring rubric. Different considerations need to be taken into account while using different approaches to its design. One of the purposes this study pursued was analysing the analytic and holistic scores that

48 EFL teachers gave to five written samples. Since the researcher developed the holistic and analytic scoring tools used, the following section gives a description of the suggested processes to follow to elaborate this tool.

2.7.5 Developing a scoring tool

Creating and adapting rubrics for classroom use are a task that language managers and language teachers should consider prior to the assessment process. Although large-scale assessment rubrics may be useful for the classroom context, Weigle (2002) considers that these may not always be the most appropriate. They should be adopted with serious consideration of the teaching and learning goals and recommends using `scoring instruments that are specific to the assignment and to the instructional focus of the class and that provide useful feedback to students´ (Ibid, p.188).

Distinct factors should be considered when developing a scoring rubric, such as the people that will use the scores, the aspects of the text, the points and scoring levels that are to be included in the rubric and the mode in which the scores will be reported (Weigle, 2002). Additionally, it is necessary for the tool to be useable and interpreted easily by scorers, or any other agents that have a role in the assessment and evaluation process. When adapting, or constructing it, the function of the tool depending on who is going to use it, needs to be considered. For instance, a) constructor-oriented scales are meant to guide the assembly of the task and refer to the kinds of writing test takers are expected to encounter in a specific level of the scale (Bachman and Palmer, 2010; Weigle, 2002); b) assessor-oriented scales are designed with the purpose of assisting the rater in the scoring process and comparing the text with the descriptors on the scale; and finally c) user-oriented scales have the purpose of guiding the user of the test to the interpretation of the scores and provide useful

information. Another important consideration is the amount of level descriptors that the scale should include.

According to Weigle (2002), the number of descriptors that should be included in the scale depends on aspects such as the range of performances expected from the test takers as well as the experience and background of raters. The more experienced the raters are, the easiest it may be for them to distinguish among multiple levels of performance thus less experienced raters may obtain more reliability with less descriptors on the scale. In relation to this aspect, Bachman and Palmer (2010) state that it is necessary to consider the number of distinctions raters can reasonably make consistently and add,

> It would be easy to create rating scales, to say, ten ability levels but it is unlikely that raters could make so many distinctions with any kind of consistency. The program director will also need to consider the meaningfulness of the scales, in terms of the degree to which they correspond to levels of ability that are assumed for the different levels of the course (p.342).

In their contribution, Bachman and Palmer (2010) consider the importance of meaningfulness, practicality and consistency in the elaboration of scales for writing assessment that have placement purposes. It is put forward their interest in developing scales that focus only on language skills without considering the context or specifications of the test environment. Finally, they add that the definition of the scale should include the description of the language features to be rated and the description of the degree of mastery of each feature included (Ibid, p.343). Abdul Raof (2002, Abdul Raof *et.,* 2011) adds to Weigle´s (2002) and Bachman and Palmer´s (2010) suggestions that four to five levels of descriptors for rubrics are enough for raters to provide a valid interpretation of language performance while highlighting the importance of including linguist specialists in the construction of a rubric as well as workplace stakeholders (Ibid, 2015).

For this study, the main researcher took into consideration the meaningfulness that the 5 language categories and the 5-point scale could have for EFL language teachers in the Mexican university context. Finally, the possible consistency with which inexperienced participating teachers would use the rubric was also accounted for. By adapting the language used to describe each category and making its use as easy going as possible, it was intended to seek for as much consistency as possible. In general, writing assessment and scoring procedures represent a difficult task during which language instructors may face specific issues and problems that could jeopardize the reliability of students´ assessment. However, it is necessary to care for these factors in order to care for the importance language assessment has for students. The next section focuses on describing the important role of assessment in language development.

2.8 EFL/ESL Writing Assessment Issues

Some of the issues language teachers and large-scale test raters face when assessing writing are related to the judgements or assessment performance of the scorers. Since writing assessment depends on human judgement, score inconsistency among writing samples is always present (Bachman and Palmer, 2010; Hamp-Lyons, 1990; Weigle, 2002). Variables such as academic background or assessment experience may have an important role in the consistency of scores. On the other hand, the different interpretation that scorers may give to the scoring scale or rubric may also represent a crucial inconsistency factor. Scorers may interpret the scale with different levels of severity, they may take a long time to understand it, to get accustomed and familiarized with the scoring scale or they may give more importance to aspects in the text that are not actually included in the scoring scale (Bachman and Palmer, 2010, p.353). In regard to intra rater reliability issues such as scorer fatigue, amount of papers to score, illness, time of day, scorer

affective mood or other personal factors may also have influence in the score provided. Researchers, such as the ones discussed below, have conducted studies that have intended to understand these issues and explore possible solutions.

2.8.1 Scorer Issues

Different aspects that have been found to complicate the reliability of writing assessment is the specific background of assessors. Rater background literature is one of the most developed topics. Studies such as those lead by Barkaoui (2011), Esfandari and Myford (2013), and Lim (2011) describe how distinct background of raters influence or not their rating behaviour and their actual scores and scoring procedures.

Barkaoui (2011), for instance, focused on 31 novice ESL teachers and 29 expert teachers. The larger group of teachers had minimum five years of teaching and rating ESL writing and had MA or MED degrees while the smaller group were teachers who had completed a pre-service or teacher training program. Participants rated analytically and holistically 24 ESL papers each. Data was analysed with FACETS computer software program with the purpose of estimating test-taker writing ability, inter-rater agreement, and rater severity and self-consistency (p.282). Data found suggested that raters were less severe during analytically scored papers and it allowed obtaining greater intra-rater reliability while holistic assessment led to greater intra- rater reliability. It is concluded that analytic scores may be more useful if details about student writing are provided and that both assessment methods are quite necessary in different contexts and may be useful with different assessment purposes. Finally, the importance of rater training is pointed out as a means of obtaining greater reliability. Although this study reflects the usefulness of both scales, it does not reflect scorer´ assessment process considering the actual involvement of teacher

assessment literacy or assessment training. Other variables that may have been analysed could include the nature of the text assessed and its level of difficulty.

Esfandiari and Myford (2013) contribute to rater type research by presenting the results of a comparison of three different types of assessors and their severity when rating. Participants included 188 university students enrolled in a public university in Iran and six male university teachers with an English Language Teaching academic background. All participants were provided with assessment training. Student participants were provided with guidance to carry out self-assessment and peer-assessment while teacher participants were provided with individual training for rating. After training, the students assessed their own paper and each teacher rated 188 papers each with an analytic rubric over a month. Data was analysed using many-facet Rasch measurement model and it was found that peer and teacher-assessors were more severe than self- assessors. Although training was provided, researchers concluded that allowing more time for information of training to `settle in´ (p.126) may imply a difference in self-assessors´ leniency. Interesting is to notice that this study heavily relied on quantitative methodology in which scorers answered surveys and scores provided to papers were analysed. It would be worthwhile considering a qualitative approach in which a more open- structured analysis of data could provide a more detailed and in-depth perspective of stakeholders´ involvement in assessment.

Lim (2011) in a longitudinal study described how experienced and inexperienced raters scored the writing section of a proficiency test over 3 periods of 12-21 months. The analysis of ratings focused on rater severity and consistency by using Multi-Facets Rasch Analysis on the FACETS program. Results may suggest that, although their ratings had worse quality in comparison to more experienced peers, novice teachers had the ability to

moderate their ratings quite easily and learned how to rate fairly quickly therefore being able to improve and maintain rating quality. It is also suggested that raters could be consistent in their reliability among the three periods of time without undergoing any further training therefore affirming the existence of experienced raters. The author concludes by suggesting that a strong relationship among the quantity of ratings and the quality of ratings exists. Different types of rubrics have become important assessment tools among teachers and important factors that also impact the assessment of writing. The following section describes studies that compare scoring tools and the influence that these have on writing assessment.

2.8.2 Scoring Scale Use Issues

Researchers such as Barkaoui (2007) and Knoch (2009) have analysed how distinct rubrics, mainly analytic and holistic rubrics, make a difference in raters´ scoring. Barkoui (2007) compared the impact that the use of two different rubrics (multiple-trait and holistic) had on the ratings provided by four Tunisian EFL teachers to thirty-two argumentative papers. From these papers, twenty-four were rated silently and eight were rated while engaging in think out loud protocols. By following a Generalizability Theory (G-) approach, variance of scores was analysed considering variables such as the interaction of students, topics, raters, and the holistic and analytic rating scale used. Verbal protocols were transcribed and separated into decision-making statements assigning each statement a code. Data found suggested that holistic rubrics result in better score reliability while multiple-trait scores resulted in higher variability among scores. Verbal protocols revealed that raters relied heavily on the impressionistic criteria instructors used in their regular assessment practice and researcher points out that although supplementary criteria were used to assess writing it led to the higher consistency of holistic scores. Finally, the researcher points out the usefulness of training to improve the use of rubrics and improve

score consistency. Verbal-protocols may be a subjective data collection tool to use since it requires from the participant to focus in the task that is being developed and reflect on the reasons for conducting such a task (Dörnyei, 2007). Therefore, other collection tools such as interviews or assessment observations may provide the wanted detailed response allowing the participant focus on a single task.

With the intention of approaching the importance of scoring scales for diagnostic assessment in EAP contexts, Knoch (2009) compared two rating scales (one newly developed and more descriptive analytic scale and another commonly used analytic scale) and their use among ten trained raters on one hundred papers of a large-scale diagnostic assessment test used in a public university of Australia. Raters were provided with training to use the rating scale and a manual for home-study previous to training. Data collection instruments included questionnaires and interviews (seven raters were interviewed) to account for the perceptions of raters in relation to the efficiency of the scales. Scores were analysed with multi-faceted Rasch analysis while transcripts of interviews were analysed separately. Results found indicated that when comparing individual traits of both scales, smaller differences were found between raters´ score leniency and harshness as well as higher rater reliability. When compared as a whole, raters considered more aspects of the text but it also led to greater score inconsistency. In terms of rater perceptions, participants stated to assign a holistic score on an analytic trait while using the old rubric due to the ambiguity of the descriptors. However, while using the new rubric raters felt more comfortable because descriptors were considered to be more explicit. The researcher pointed out the importance of descriptor explicitness among the same type of rubrics and emphasized the usefulness of context specific standards and rubrics.

To determine the usefulness of a rubric, Saxton *et al.* (2012) focused on the effectiveness of a context-specific analytic scoring rubric (The Critical Thinking Analytic Rubric) that emphasizes the importance of critical thinking skills in high school English natives and L2 writers and tried to prove its effectiveness in their specific context. Two trained raters with similar backgrounds scored samples in two distinct time periods: 2008-2009 and 2009-2010. One hundred and fourteen papers in the first period and thirty randomly selected papers in the second, were scored with the same rubric. Raters received training prior to the scoring process. Data was analysed using the consistency measure, Cronbach's alpha, with the SPSS software program. Researchers found that in terms of intra-rater reliability levels were considered acceptable. Inter-rater consistency levels were also found to be acceptable and it was considered that the two raters, with the appropriate training, were capable of demonstrating acceptable levels of consistency on the samples of both periods of the study.

2.9 Chapter Summary

This Chapter focused on describing writing as a language skill and how its assessment should be perceived as a social activity (Scarino, 2013). It described assessment as an activity that largely depends on the environment in which it evolves and the context in which learning takes place. Therefore, suggesting that people involved in this context or their personal characteristics may have an important role in the outcome of assessment.

This project seeks to analyse how a suggested solution to obtain higher degrees of reliability, assessment training, can have an impact on classroom assessment of writing. This leads to the question if writing assessment training provided to EFL teachers makes a difference in their assessment process in the classroom and in obtaining higher levels of

reliability of scores. Experts have come to suggest that raters and teachers should be trained in how to use the specific scoring scales that are to be used so that differences in the language proficiency are spotted more easily and therefore scored more accurately (Hamp-Lyons, 2003; Bachman and Palmer, 2010; Hamp-Lyons, 2003). It is also suggested that training may help improve the processes teachers follow in their classrooms. Considering this suggestion and the major importance that assessment literacy has to this project, the following chapter focuses on the concept of assessment literacy as an important factor that may lead to the improvement of EFL teacher assessment practices.

Chapter 3: Assessment Literacy and EFL Writing Assessment

This chapter addresses the concept of assessment literacy from a social dimension. It considers that providing teachers with the knowledge of assessment is not enough to ensure valid and reliable assessment of students´ language skills. But instead it is also necessary to connect teachers´ knowledge with classroom assessment and large-scale testing (Fulcher, 2012). Additionally, this chapter seeks to provide an analysis of the different areas approached by assessment literacy research seeking to identify the area of contribution of this study.

3.1 The Nature of Assessment Literacy

Assessment literacy is considered a social practice (O´Loughlin, 2013) that is gaining importance in different areas of education. Scarino (2013) and Moss (1996, 2004) also view assessment literacy from a sociocultural perspective in which the knowledge of assessment and its practices rely on the interpretation of the assessor and the preconceptions that are brought to the process. Interpretation is involved `…in conceptualizing the construct, in considering tasks that are intended to elicit students´ performances and the way these tasks are interpreted by students, in interpreting and applying criteria and standards for judging performance and in interpreting evidence as part of processes of validation´ (Scarino, 2013, p.323). Additionally, preconceptions such as previous experiences, assessment traditions or assessment context may also be involved in the building the assessment process.

According to Fulcher (2012), the rise of the concept is a result of three major changes in assessment and testing. The first is the extensive use of large-scale test results for policy making and for keeping control over teachers' performance. The second is the role that tests have in the globalization of languages. Immigration policies in developed countries around the globe have had a need to guard their cultural identity thus considering language part of that identity, resulting in language test preparation instruction important for policy makers. Although teachers are not directly affected by test results, their instruction is. It focuses on meeting language testers expectations on tests´ results for international mobility and the financial value of tests. Finally, the third factor was the need for teachers to have a range of strategies to implement assessment and to evaluate the degree of success. In terms of assessment and evaluation, `language teachers are expected to choose or construct, administer and interpret the results of assessments designed for a variety of purposes and situations´ (Stoynoff and Coombe, 2012, p.122). However, assessment and evaluation are not easy processes and when conducted without valid and reliable procedures, consequences on students´ and teachers´ performance may be jeopardized.

The importance of assessment literacy lies in that `without a higher level of teacher assessment literacy, we (teachers) will be unable to help students attain higher levels of academic achievement´ (Coombe *et al.*, 2012, p.20). Therefore, the need for teachers to obtain academic preparation in language assessment is of major importance.

Assessment literacy is a concept, which was developed as a result of the increased demand and use of assessment data by involved stakeholders (Inbar-Lourie, 2013). Most of these are inexperienced in the language teaching field and have little or no knowledge of how and what to assess. Inbar-Lourie (2008, 2013) and Malone (2011) consider that the term

assessment literacy refers to the knowledge that instructors may have and how they choose to use this knowledge in their assessment practices. Taylor (2012) agrees by adding that assessment literacy refers to teachers´ familiarity with measurement practices and how this knowledge is applied in the classroom when assessing language. Inbar-Lourie (2008) specifies that while assessment is context specific, the concept of assessment literacy implies a social factor in which teachers, assessment techniques and evaluations are embedded in a specific social situation. Other researchers believe that any course or training that seeks to enhance instructors´ assessment literacy needs to consider the assessment culture of each program and institution (Taylor, 2009), thus suggesting that assessment is a `social practice and a social product ´ (Inbar-Lourie, 2008, p.387; Fulcher, 2012). Scarino (2013) adds that teachers need to obtain an understanding of their own knowledge, values, conceptualizations, interpretations, judgements, decisions, and experiences (preconceptions) to give way to new knowledge about assessment through self-awareness as a teacher and an assessor. It is considered that the ultimate goal of assessment literacy is the modification of knowledge and understanding through the experience of new preconceptions (Ibid, p.324). A teacher who is assessment literate may have a wide repertoire of assessment tools as well as techniques and is capable of deciding which technique is appropriate in a specific circumstance (Coombe *et al.*, 2012). However, I believe that assessment literacy goes beyond the knowledge of assessment, its practice in the classroom, or the involvement of social and contextual factors. It may also involve the process in which teachers reflect on their own assessment practices in the classroom as a means to reconceptualise their assessment knowledge and practice. Once this reflection is rooted in the teachers´ regular practice, the will to innovate and improve their assessment may also be encouraged. This last stage, I believe is the most difficult to attain. Therefore, I would argue that assessment literacy is a crucial element of teacher education and

professionalization programs (Xu and Brown, 2016) for it may allow teachers the opportunities to engage in teacher reflection.

In this sense, experts such as Xu and Brown (2016) provide a conceptual framework of Teacher Assessment Literacy in Practice (TALiP) which portrays the combination of pre-service and in-service teacher principles, the knowledge and skills of language assessment and the specific socio-contextual aspects of assessment such as policies, cultural values and social norms (p.2). Based on an extensive search of previous assessment literacy studies, the researchers proposed the TALiP in an attempt to provide the field with a framework that not only includes theoretical knowledge of assessment and teacher sociocultural perspectives but also one that includes the development of teachers throughout their teacher education programs. As portrayed in Figure 6 below, the framework consists of seven levels of assessment knowledge, skills and practice. The levels begin with the a) knowledge base level, then moves upward in the pyramid to b) the interpretative and guiding framework, c) teacher conceptions about assessment, d) macro socio-cultural and micro institutional contexts, e) teacher assessment literacy in practice, f) teacher learning and g) assessor identity (re) construction (Xu and Brown, 2016, p.19).

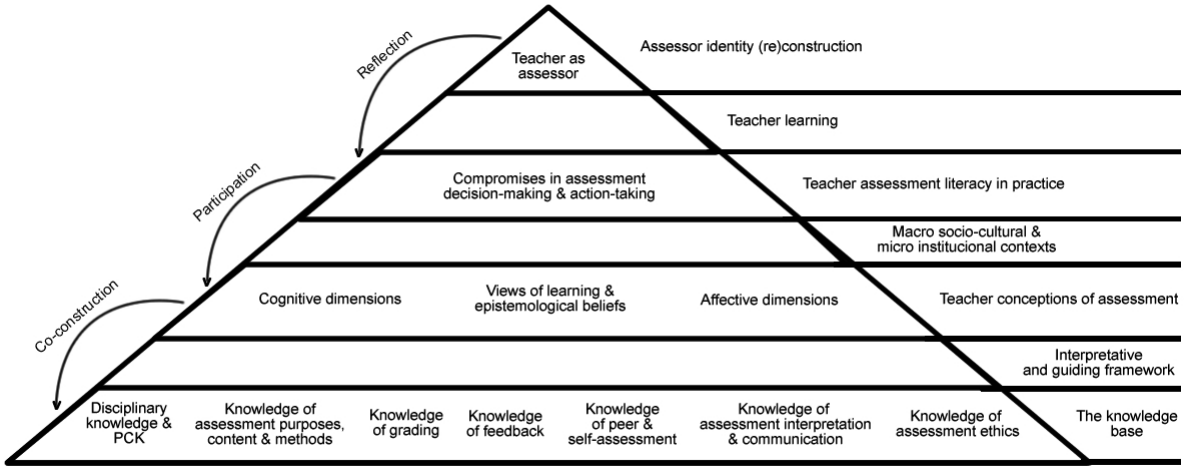

Figure 6 Teacher Assessment Literacy in Practice (Xu and Brown, 2016, p.19)

The authors conclude by arguing that the top level of the pyramid `Teacher assessor (re)construction´ is the ultimate goal of assessment literacy since it is through the constant questioning and reflection of teacher knowledge and classroom experience that they may find a change in their practice. It is my belief that although the model encompasses stages that allow teachers to gain literacy in assessment and encourage improvement in their assessment procedures, it is a model that requires teachers´ time investment and possibly money investment to complete. Since the main premise is that the TALiP is mostly possible with teacher education as a cyclical model that requires constant teacher reflection, educating teachers requires a lot of time and financial resources. This, I believe, could be a possible explanation of teachers dropping out of the assessment literacy challenge.

Providing teachers with this literacy has represented several issues in countries such as the United States. For instance, the lack of assessment courses in TESOL undergraduate and graduate programs and/or the lack of teacher-students enrolled in those programs that do offer the course; and finally, the fact that current language teaching practices do not provide assessment the importance it needs therefore losing connection between classroom and assessment practice (Crusan, 2014; Stoynoff and Coombe, 2012; Weigle, 2007).

In the Mexican context, EFL teachers may lack the necessary knowledge to develop their classroom assessment practices while language institutions occasionally take this `teacher responsibility´ for granted (López Mendoza and Bernal Arandia, 2009; Lam, 2015). It is assumed that teachers should already know how to carry out since it is conducted on weekly, monthly or bimonthly basis. However, EFL teachers occasionally feel they do not

have the necessary professional skills to assess writing (Metler, 2003) thus avoid its assessment or even teaching it in their classrooms.

3.2 Perceptions of Assessment Literacy among Stakeholders

According to Coombe, *et al.* (2012) assessment literacy divides its area of study in two sections: a) what teachers know about assessment and b) what teachers perceive of their knowledge of assessment. In relation to the views and perceptions, these authors state three stances: that of the student, the teacher and the educational boards which establish curriculum policies for language programs. While students face anxiety and fear when being assessed, teachers do not come upon a very different landscape. Teachers who are not involved in the elaboration of the tests or assessment tasks often feel that the connection between the classroom and the actual test is lost. Finally, in terms of educational boards the authors state that,

> …in the field of English Language Teaching, TESOL partnered with the National Council for the Accreditation of Teacher Education (NCATE), created the TESOL/NCATE standards for ESOL teacher Education. Assessment constitutes one of the five knowledge domains within these standards. In Europe, the Common European Framework of Reference and the European Portfolio for Modern Languages are requiring teachers to adopt new ways of assessing language ability (Ibid, p. 21).

With the creation of these teacher education standards and the inclusion of assessment as a component of language teacher education, educational boards are acknowledging the importance of having teachers develop professionally their skills in the area of assessment. If assessment instructors are assessment-literate then identifying appropriate assessment for specific purposes (student placement or diagnosis) and analysing data to decide how to improve teaching and instruction may be possible (Coombe *et al.*, 2012).

The case of assessment literacy in foreign language contexts paints a different picture. For instance, in countries such as Mexico the National Institute of the Evaluation of the Education (Instituto Nacional para la Evaluación de la Educación) has the purpose of assessing public education, teachers who are in service in public Mexican education, and their teaching competencies to assure quality and fairness education for students (INEE, 2013). However, this board does not consider the evaluation of language teachers´ competencies and/or teachers´ language assessment skills in their education evaluations since it is out of the scope of their objectives. However, it is interesting to point out that up until the writing of this thesis the Ministry of Education in Mexico (Secretaría de Educación Pública) does not include a language assessment component for their English teachers to study in their teacher professionalization programs. Similarly, Falvey and Cheng (1995 cited in Coombe *et al.*, 2012) and Lam (2015) pointed out that teachers in Hong Kong believe they received little or no training at all in assessment while teachers in Israel and Colombia felt they did not have the sufficient knowledge and training to carry out assessment procedures in their classrooms (Coombe *et al.*, 2012; Shohamy *et al.*, 2008; Lopez Mendoza and Bernal Arandia, 2009). Therefore, adding to my argument that in Latin American contexts where English is taught as a foreign language and where assessment literacy is emerging, educational boards still do not provide sufficient opportunities for teachers to become assessment literate therefore diminishing the importance it has for student assessment.

In Lam´s (2015) study, focus groups with pre-teacher students and interviews with instructors indicated that more than 50% of students stated to have not heard the term `assessment literacy´ while one instructor indicated that her assessment students preferred

to learn how to create tests rather than the principles and practice of assessing for learning. This may suggest that students and teachers do not have enough access to the ` knowledge base of the social dimensions of language assessment such as validity, fairness and test impact on teaching and learning ´ (Lam, 2015, p.183). Finally, the researcher pointed out that students and instructors equally considered that the content of assessment courses would not be useful to teachers since most schools still have traditional ways of assessing language.

3.3 Issues faced in Assessment Literacy

There are several factors that may hinder assessment literacy among teachers. Firstly, Malone (2008, 2013) considered that although standardized language education, teacher certification and the number of administered tests had increased there still did not exist a consensus on what teachers are required or need to assess language proficiency. For instance, do teachers need specific experience, practice or training to assess their students? In terms of writing, what do scorers need to comply with regarding their professional background and assessment procedure knowledge to rate their students´ work? These and other factors are still unclear making the assessment of writing more subjective than it may already be and the consistency factor more difficult to improve.

Another issue pointed out (Taylor, 2012; Jeong, 2013) is the condition in which assessment courses are provided to language teachers. Assessment literacy is generally provided to teachers in either undergraduate or graduate courses during which contents and structure are undefined and generally provided by instructors who are not assessment literate themselves. Thus, being unclear as to how these courses should be provided or what should be considered in their contents. For instance,

> …is there a required level of assessment literacy to teach an LAC? (Language Assessment Course)...do the instructors who teach language assessment courses have a testing background? What kind of assessment knowledge do they possess? How do they teach the course and does it meet the expectations and needs of the student teachers? (Jeong, 2013, p.346)

This is an issue that plays a key role in reaching teacher assessment literacy. It is necessary for language instructors who assess their students and make decisions based on these assessments to have sufficient knowledge and be able to reflect on when and where to use that specific knowledge (Coombe *et al.*, 2012).

On the other hand, fear among language teachers represents one of the major barriers to assessment literacy (Coombe *et al.*, 2012; Stiggins, 1995). Many teachers accumulate negative perceptions, such as fear, anxiety and dislike (Crusan, 2014) towards assessment, which do not allow them to be open to reflect on their own assessment competences. Other reasons include daily teacher workload (it is easier to let others deal with the assessment part of teaching) little time available to dedicate to assessment and the lack of administrative financial resources targeted to the literacy of assessment.

Despite these constraints and because assessment is part of a teacher's everyday practice, it is of major importance that they learn about assessment and become involved in the creation and administration of tests at their workplaces (Crusan, 2014). The following section focuses on writing assessment and the literacy that writing teachers may need to carry out their regular classroom assessment.

3.4 Assessment Literacy in practice: Training Language Teachers

It has been suggested that issues in the assessment of writing such as inconsistency of scores, scorer or examiner fluctuation may be diminished with the use of detailed scoring rubrics and assessment training (Fulcher and Davidson, 2007; Weigle, 1994). According to Weigle (2007) and Crusan (2014) for teachers to be capable of assessing writing they need to a) understand distinct assessment methods and choose the corresponding method according to their specific assessment objectives; b) be capable of recognizing good writing and good assessment; c) comprehend important concepts such as summative and formative assessment, validity, reliability and practicality; d) understand and put into practice the test development process; e) create and use assessment tools that are effective in attaining the assessment objective; and finally f) be aware of external large-scale tests, their purpose and the interpretation of scores.

For these traits to be obtained among trainees, experts have suggested several factors to consider when providing training to language instructors. For instance, Bachman and Palmer (2010) consider that selection of trainers and training content have an influence in the training of teachers. When selecting teachers, it is vital to consider the language ability of the teacher in relation to the written texts they will assess. In other words, if teachers will assess a text with a high level of language mastery then they should have a complete mastery of the language and vice versa for less proficient texts.

Content of training sessions may consider the following stages (Bachman and Palmer, 2010, p.p.353-354), a) read and discuss scales together, b) review language samples which have been previously rated by expert scorers and discuss the ratings given, c) practice rating a different set of language samples. Then d) compare the ratings with those of

experienced raters, e) discuss the ratings and how the criteria were applied, f) score additional language samples and discuss, g) each trainee scores the same set of samples and discuss and h) selection of scorers who are able to provide reliable and efficient ratings. In a training session that follows this previous outline of steps, trainees are led into a series of activities, are introduced to the test or task as well as its purpose and format. Then, they are exposed to the specific training tool with the purpose of unifying (as much as possible) opinions and interpretations of the tool. During this training, assessors are provided with a space to discuss difficult or unclear cases in which the written text is found among the borderline of the scale. However, this training procedure is not always as linear as outlined by Bachman and Palmer (2010). Specific contextual aspects need to be considered before providing assessment training to teachers. For instance, aspects such as their assessment needs, the type of students they work with, the assessment policies of their workplaces, among others. Understanding these contextual elements will allow trainers to understand which steps of the training process need to be adapted to suit teachers´ characteristics.

Weigle (2002, p.130) considers that `…specific circumstances will dictate to what extent this exact process (training outline) can be followed…´ (p.130). She adds that assessment training should also include the type of assessment tasks teachers need to score. In doing so, trainers could include scripts that focus on different issues such as papers that do not address the task or that are found to be on the borderlines of two descriptors on the rubric. This could help teachers become aware of what to do when encountering these circumstances. Finally, Weigle (2002) goes on to suggest that it is important to communicate to teachers being trained that 100% of reliability or consistency among

scores to writing is impossible. Instead, more assessment training can be considered to diminish score inconsistency.

Research such as that carried out by Weigle (1994) or Shohamy *et al.* (1992) focus on the effects that rater training may have on assessors and the scores provided in second language assessment contexts. Others such as Nier, Donnovan and Malone (2013) focus on assessment literacy and online training for foreign language teachers. These and other experts have approached writing assessment from different angles in their research with the purpose of enlightening the path of those in the assessment field or to describe and give solution to specific assessment issues. These and other researchers have tried to provide some insight of the importance of assessment literacy for English teachers and for their assessment practice.

Previous research that has been developed in the areas of assessment literacy, assessment training for language writing teachers, rater training, intra and inter rater reliability in writing assessment can aid in establishing the basis for this project. It is important to point out that writing assessment research has focused mostly on large-scale standardized testing contexts in which a rater interacts with the assessed text and uses rubrics to provide a score. Other studies focus on writing assessment from a classroom perspective in contexts where assessment plays a summative and formative role in the classroom (Yorke, 2003). However, this project intends to aboard the rather difficult practice that tertiary EFL teachers carry out on a daily basis and the effects that assessment literacy may have on these practices.

3.5 Assessment Literacy in L1 contexts

Assessment literacy is still quite young in regard to the amount of research carried out. It is an area that began flourishing in the late 90s with studies such as those lead by Stiggins (1995, 1999) who defined assessment literacy and the importance of assessment literacy not only for the educational field but also for other fields in which it is not enough to rank students by their achievement but to also bring forward the importance of investing in teacher preparation. He notes that assessment literate teachers know what to assess, the reasons of their assessment, the best technique/method to approach skill assessment, the best way to obtain performance from students, the potential issues with assessment, and solution to these issues. But above all assessment literate teachers are aware of the potential consequences of inaccurate assessment.

Other studies such as those lead by Volante and Fazio (2007), Metler (2003) and Metler and Campbell (2005) seek to understand assessment literacy from teachers´, teacher candidates´ and students´ perspectives. For instance, in the Canadian L1 education context Volante and Fazio (2007) focus on the analysis of the perceived assessment literacy levels of primary and junior teacher candidates. They also approach the candidate´s purposes of assessment, their use of different assessment techniques, their need for training and the promotion of assessment literacy. Sixty-nine primary/junior teacher candidates answered a nine-item questionnaire (four closed-ended and five open-ended questions). Quantitative and qualitative analysis of data suggested that pre-service teachers perceive themselves as non-competent assessors, they primarily viewed the purpose of assessment as summative and believed they needed more training in classroom assessment and evaluation.

Also, focusing on L1 contexts, Metler (2003) surveyed pre-service and in-service teachers with the purpose of understanding their assessment literacy levels and further compares the differences among these groups. The pre-service group was composed of 67 undergraduate students majoring in secondary education while the in-service group consisted of 197 teachers representing every district in a three-county area in the US. The participants answered the Classroom Assessment Literacy Inventory adapted from the Teacher Assessment Literacy Questionnaire (Plake, 1993 cited in Metler, 2003) and which consisted of thirty-five content based items, which corresponded to the Standards for Teacher Competence in the Educational Assessment of students. After conducting quantitative analysis on the data obtained, the researcher concluded that from a holistic perspective the standard that received the highest means of performance was Communicating Assessment Results while the one that received the lowest performance was Developing Valid Grading Procedures. When comparing both groups, it was found that in-service teachers scored higher than their pre-service counterparts suggesting that additional to the assessment courses and assessment training teachers are provided in their work centres or during their undergraduate studies, assessment literacy is also supported with an `on-the-job´ type of training (Metler, 2003, p.23) in which teachers learn and become assessment literate with their everyday practice (Vogt and Tsagari, 2014). The researcher concludes that further research is needed to understand the role that everyday experience may or may not add to teachers´ assessment literacy.

In a similar study and with the purpose of developing a distinct instrument to measure teachers´ assessment literacy, Metler and Campell (2005) analysed the assessment literacy levels of classroom teachers in the American context that could provide an insight into what teachers actually did in their classroom and if they knew how to do it. Therefore,

these researchers set out to develop and evaluate the psychometric properties of the Assessment Literacy Inventory which `was designed to parallel existing Standards for Teacher Competence in the Educational Assessment of Students´ (Ibid, p.2). The study included a development phase, a piloting phase and a validation phase with the participation of a total of 401 pre-service teachers in two data collection phases. The Inventory included thirty-five items that reflected classroom scenarios and then followed by seven multiple choice items. The researchers conclude that the overall reliability levels of the Inventory items were just about satisfactory (.74) thus suggesting that it may be a tool that may enlighten the path of school districts and policy makers to allocate resources to enhance assessment literacy opportunities in areas in which they are most needed. Finally, it was added that pre-service teachers scored relatively low in the inventory items (68% of items answered correctly) thus suggesting that undergraduate assessment courses may not be enough to obtain the necessary knowledge and skills to assess students´ work therefore echoing the conclusions provided by Metler (2003) in which it is considered that day-to-day classroom experience may be necessary to successfully obtain teacher assessment literacy.

Focusing on the Mexican elementary students and their L1 writing skills, Contreras, González and Urias (2009) focused on raters in the Mexican elementary school context. Participants were 31 experienced raters that took a two-session rating training course. Raters were advanced students or graduated students from the undergraduate program Language and Literature of Hispanic America offered at the Universidad Autónoma de Baja California and Spanish professors of the Escuela Normal Estatal de Ensenada. After obtaining scores on 100 papers for each rater and performing statistical analysis to determine inter-rater reliability and Rasch analysis for variable influence, it was found that

raters tended to be lenient in their scores and that reliability was possible among scores. Researchers also found that most raters provided more strict scores in their assessment while keeping their scores within the third level of student performance. Researchers are not clear in terms of the type of tasks students wrote, the rater background, nor of the nature of the two training sessions. It can be stated that the study analysed the scores to understand their leniency and strictness without considering the role of the rater and contextual traits in which assessment could have been embedded thus isolating assessment. I believe that more contextual information needs to be included through the means of qualitative inquiry to allow a deeper understanding of the assessment outcome.

3.6 Assessment Literacy in ESL Testing Contexts

Bailey and Brown (1996) explored the components of language testing preparation programs for language teachers. Then 11 years later, the same investigation was replicated with the purpose of exploring the preparation that ESL teachers had in language testing. In both studies a survey was used with Likert-scale answer choices. The researchers conclude that more statements needed to be added to the survey used in 1996 such as Test analysis, Washback, Test bias, Testing in relationship to curriculum, Standard (cut-point) setting, Critical approaches to language testing, Language program evaluation, Classroom testing practices, Rasch analysis, Computer-based TOEFL (CBT) scores, Internet-based TOEFL (IBT) scores, Generalizability theory, Consequential validity, Values implications in validity, Multiple regression, Structural equation modelling, Analysis of variance (ANOVA), Many-faceted Rasch (FACETS) analysis, and Validity as a unitary concept. Additionally, the textbooks used in these courses and considered for the analysis of the studies changed from the study of 1996 and 2007. While in 1996, thirty-two different textbooks were analysed, only twenty-nine were considered for the 2007 study.

Researchers stated that the main reason was that out-of-print-books were considered in the first study which were not available in the second. Finally, it is concluded that the differences found between both studies can shed light on the new needs of language assessment courses and therefore guide teacher trainer and course policy makers to make the appropriate choices in terms of content and structure.

A study that approached the needs of teachers in assessment training courses is that conducted by Hasselgreen, Carlsen and Helness (2004) in which language teachers, language teacher trainers and language assessment experts answered a background questionnaire and a survey in which general professional background was obtained from the former and their needs in terms of assessment literacy was obtained from the latter. Responses from 197 participants revealed that teachers and teacher trainers had very similar needs which included creating assessment tools, use of portfolios, peer/self-assessment, interpreting results, establishing validity and reliability throughout statistics, rating student performance in productive skills among others. The assessment experts considered they needed training in creating and developing items, making assessment-based decisions, using and considering the CEFR as basis and support for the creation of tests and testing processes. Although there were some issues encountered with the number and nature of the participants (some countries were more represented in the sample than others, some teachers had more than one role being analysed) this study can shed some light in terms of the need of those novice and expert teachers that consider assessment an important component in their courses and for language managers that take decisions considering assessment results.

Fulcher (2012), with a very similar purpose to that stated by Hasselgreen *et al*. (2004), conducted a study to explore the assessment training needs of language teachers and then use the information to produce print and online materials to use for teacher training. An online survey was also used as an instrument to collect data which obtained responses from 278 language teachers of different parts of the world such as Australia and New Zealand (13.5%), North America (13.5%), South America (5.4%), the Middle East (2.7%), the Far East (16.2%), and Europe (37.8%). The researcher goes on to describe his results and explains that,

> …language teachers are very much aware of a variety of assessment needs that are not currently catered for in existing materials designed to improve assessment literacy. The answers to the constructed-response questions in particular are indicative of changes in our understanding of the role of testing in society and a desire to understand more of the `principles´ as well as the `how-to´ … (p.125).

Finally, and in agreement with Metler (2003), Fulcher (2012) adds that most important of all, the process of combining the theoretical principles of assessment with the actual practice should also be extended to link as much as possible large-scaling assessment with the actual classroom-assessment teachers approach in their everyday teaching practice.

3.6.1 Writing Assessment and Rater Training in ESL Contexts

Under the theoretical support of assessment literacy lies raters´ training needs to assess productive language skills in large scale contexts. In terms of the impact of training on writing assessment, studies such as those lead by Elder *et al*., (2005, 2007) and Weigle (1994, 1998) can be considered pillars of this branch of assessment literacy.

With the intention of analysing assessment literacy of scorers but from a university standardized testing context, Elder *et al*., (2005) focused on describing the perceptions of

the feedback provided to the scoring processes of eight experienced raters of English diagnostic writing as part of their online training. Participants answered a pre-training questionnaire, participated in an online training course in which scripts were rated for scoring practice, participated in a post-training scoring session of fifty randomized papers, then finally received group feedback and individual sessions of feedback before rescoring 62-64 papers. Data obtained from the ratings of fifty scripts were analysed with the multi FACETS software program. Results suggested that feedback was perceived as useful by participants and suggested assessors became aware of their own rating behaviour. It was also found that raters´ scores became more consistent after receiving the feedback. Finally, the researchers concluded that the factor of improving rater inter-reliability reduced test´s discriminatory power suggesting that the cost of online training and feedback sessions outweighs the benefits of this approach. It can be argued that the benefits of assessment training in relation to the financial costs and the human resources needed to train teachers may not be immediately visible since change in teachers´ assessment practices need time to develop. However, I believe that once this process is initiated the financial investment in teacher training will gradually converge as time goes by. Time is needed for teachers to reflect, process and implement new assessment knowledge in their classrooms especially because these innovations depend on many contextual factors as suggested by Scarino (2013) and other experts.

Weigle (1994), with the purpose of exploring the effects of training on experienced and inexperienced raters of ESL placement compositions in a university context, analysed the pre- and post- training ratings provided by eight inexperienced raters of ESL placement compositions and compared them to those given by eight experienced raters. Then, the researcher identified which of the raters had differences in their Pre-and Post- training

ratings of three points or more and analysed verbal protocols of these participants while rating six different papers. Results indicated that new raters were more influenced by the training than the experienced ones. Additionally, it was found that training helped raters understand scoring criteria, it allowed them to modify their expectations of student writing therefore being more objective in their assessments and to be more aware of the importance of consistency among other raters within the program. Weigle (1998) in a different study focused on the same sixteen raters (eight experienced and eight inexperienced) and their severity and consistency levels before and after taking training. Each rater assessed fifteen samples from one type of task and fifteen from a second different task prior to the training session and two sets of sixteen different essays post to the training session. Each participant took one rating session of approximately 90 minutes. The researcher analysed the scores provided to the essays by using the IRT FACETS (multi-faceted Rasch analysis) software program with the purpose of finding and comparing the severity levels between each rater and the spread found among them. Results indicated that inexperienced raters were slightly more consistent in their levels of severity after training. However, the spread of scores among all raters was quite significant indicating that despite training raters diverged significantly from one another in their severity. The researcher concludes that inexperienced raters are more severe and less consistent in their ratings than the experienced raters before training and that although training does not guarantee consistency in severity it can encourage raters to be more internally consistent. In other words, training may allow for more intra-rater reliability. Both these studies provide an important insight into how training can be of benefit for the assessment of ESL students and rater performance in large scale testing. However, it seems that much attention has been given to this context while classroom assessment has been underexplored. The needs of a classroom and of those involved in it may vary greatly.

Additionally, I consider that other data collection tools could be implemented to fully understand the assessment process. Tools such as interviews and focus group discussions in addition to scores provided to papers may provide a more in-depth analysis of the effects of training. Additionally, more updated research is needed to picture the actual impact that training may have not only on test scores in a writing-oriented course but also in a classroom in which writing is not the only skill being taught.

Knoch (2011) conducted a longitudinal study (sixteen-month period) in which the rating behaviour of nineteen raters assessing a large-scale English for Specific Purposes (ESP), specifically of health professions over eight administrations was tracked and documented. Raters were provided with feedback of their rating performance after each administration which was generated with the many-facet Rasch measurement. The researcher intended to explore if the feedback provided was perceived as useful for the rating process. Information from scored scripts, questionnaires and interviews to raters suggest that although feedback was viewed as positive, the quality of assessment did not change when raters received feedback in comparison to when it was not provided. The author concluded that there was not a direct link among views of feedback and the impact of feedback. It would be interesting to analyse the type of feedback, its content, its focus as well as the modes used to provide feedback. It is my belief that a variety of delivery modes may provide different results in terms of the usefulness of feedback and its positive or negative impact.

3.7 Assessment Literacy in EFL Contexts

In foreign language contexts, assessment training has been analysed in relation to teachers´ perceptions of training courses (Lopez Mendoza and Bernal Arandia, 2009; Nier *et al*.,

2013; Malone, 2013; Jeong, 2013) and the needs that they consider should be covered in an assessment training course. Other studies focus on the analysis of specific assessment courses and their contribution to the assessment literacy of a teacher or group of teachers (Koh *et. al*, 2017; Lam, 2015).

With the purpose of understanding the perceptions of eighty-two Colombian EFL teachers, Lopez Mendoza and Bernal Arandia (2009) developed a survey that elicited from participants their general background and their views in regard to assessment, their use of assessment in the classroom, their scoring of assessment and the provision of feedback to their students. Researchers found that trained teachers had more positive views towards assessment in comparison to the non-trained teachers. The latter considered assessment a tool to monitor learning, to communicate with the student, to align learning and teaching and finally to empower students while the former considered assessment mandate, a summative process and a tool of power and control over students. It is noted that the majority of the teachers use traditional methods of summative assessment while a small percentage use more authentic methods of assessment. It is concluded that a correlation among teachers´ previous assessment training experience may have a role in their use and perceptions of assessment.

Koh *et al.´s* (2017) longitudinal study conducted in Singapore, analysed twelve Chinese teachers´ assessment literacy regarding the quality of task design. It did so by considering their enrolment in a two-year professional development program, which focused on developing teachers´ design and use of authentic assessment tasks. After each phase of professional development (four sessions in total), teachers designed samples of assessment tasks. Then, teachers were taught to provide each other feedback and to judge sample

student responses to their tasks designed. Data obtained from the designed tasks, the sample answers provided by students and the scores provided by teachers to these tasks suggested that after two years of professional development inter rater reliability among assessment was satisfactory while sustained improvement on teachers´ task design was shown by the end of the first year of professional development. The researchers conclude that participants strongly relied on the elicitation of linguistic procedural knowledge in the tasks they designed even though the professional development offered, strongly emphasized the importance of integrated skills tasks.

Jeong (2013), on the other hand, focused on 140 instructors that completed an online survey and six language testers that participated in a semi-structured interview. The study had the purpose of analysing how Language Assessment Courses (LAC) are constructed and taught by instructors in different countries. The researcher found that most of the student teachers (two thirds) were enrolled in language assessment courses that were taught by an instructor who was not assessment literate or language testing literate (according to the researcher´s classification). The instructors spent most of their instruction time on topics such as test theory, classroom assessment, alternative performance assessment, test specifications, and rubric development. Therefore, it was concluded that the ultimate outcome of assessment teacher training courses will largely depend on the academic background and the personality of the instructor even if the structure is similar or the same. Finally, it is pointed out that although all six language testers were aware that assessment courses were provided by professionals that were not experts in the field of assessment, this was a necessary road of action as a result of the lack of teaching staff.

With a different overall purpose Nier *et al*. (2013) and Malone (2013) focused on analysing online assessment tutorial materials and its usefulness to EFL teachers in the United States. After data analysis was obtained from the answers of eighty EFL teachers to an online survey, Nier *et al*. (2013) concluded that most of the foreign language teacher participants considered online training useful for their future assessment practice but more examples and samples were needed to further understand the process of assessment. It was also found that online tutorials allowed teachers to feel more comfortable with specific assessment terms. Malone (2013), adds to this previous study, the analysis of an online assessment tutorial from the perspective of language experts and foreign language teachers in the United States. After seventeen language testing experts and forty-four language teaching experts participated in focus group interviews and answered an online survey, the researchers found that language testing experts considered important the fidelity of testing definitions and appropriate test use be cared for in online resources such as the online tutorial under analysis. On the other hand, the language-teaching experts considered aspects of presentation and delivery of materials. It was interesting to note the differences in the focus of feedback of both groups of reviewers as well as the questions that the researcher arises as an outcome of this research. For instance (Malone, 2013 p.342),

> how can resource developers combine fidelity of definition with succinctness, particularly given the often nuanced and technical nature of language testing definitions? In developing such resources, how can precision be balanced with clarity? Who should review such resources and who should determine how much technicality is sufficient for language instructors?

This study adds to those previously described in regard to how language teaching experts/teachers perceived existing assessment literacy material and its effectiveness in their actual assessment practice. Finally, it is important to bring forward that the perceptions and assessment needs of language teachers may be very different to those of

language testing experts (Malone, 2013). Therefore, a sharing-point needs to be met so that both views nourish each other.

With a very similar purpose, Lam (2015) explored the status of assessment literacy among tertiary teacher education programs. Secondly, he analysed the extent to which two assessment courses encouraged or not the assessment literacy of pre-service teachers in one teacher education institution. Data obtained from document analysis, focus groups with pre-service teachers, interviews with instructors and surveys suggest that there is not enough promotion of assessment literacy among teacher education institutions in addition to the fact that they do not provide sufficient assessment courses to equip pre-service teachers with assessment strategies during their studies. However, it was found that five twelve-hour courses were offered between 2013-2014 for in-service teachers only. Therefore, the researchers concluded that assessment literacy courses in Hong Kong were not enough to satisfy the needs of pre-service and in-service teachers.

Vogt and Tsagari (2014) attempted to investigate the perceptions of foreign language (FL) teachers in Europe in regard to their experience with assessment training, and their need to be trained in different areas of language assessment. Specifically, teachers were teaching in primary, secondary and tertiary levels in the European countries of Cyprus, former Yugoslavian Republic of Macedonia, Germany, Greece, Italy, Poland and Turkey. The researchers used a mixed method approach in which data were obtained from surveys and semi structured interviews. Data revealed that the area that needed to be the most reinforced among teachers was `purposes of testing´ while 42.4% of the surveyed teachers claimed to have not received any training at all. Vogt and Tsagari (2014) conclude that assessment procedures such as designing tests, giving grades, placing students in their

corresponding levels, and awarding certificates are not fully developed skills in teacher participants and most probably they are learned on the day to day practice. It is pointed out that most of the teachers preferred advanced training to improve their assessment practice and they perceived the need to have further training in assessing productive and receptive language skills, micro linguistic aspects, the assessment of integrated skills and statistical analysis for language assessment (Hasselgreen *et al*., 2004; Vogt and Tsagari, 2014). Teachers reported to feel prepared to design and develop tests that correspond to traditional forms of assessment and they compensate for the lack of proper assessment training by learning on the job (by observing a mentor or other colleagues). The results of this study possibly mirror the Mexican context in which teachers need to develop their skills to create assessment instruments. However, assessment training in our context still needs to address the issue of teachers´ overuse of traditional forms of assessment. Exploring alternative assessment instruments may also allow teachers have additional tools to assess students as well as become more confident in their procedures.

This section has attempted to identify the main studies carried out in two different contexts that may entail differences for teachers, students and other stakeholders involved in the language assessment process: ESL and EFL contexts. In an ESL context, a student may be enrolled in a course that focuses on a specific skill, for instance a writing course, in a university setting after which students need to provide proof of their English proficiency to be enrolled in their university majors. They are involved in a context in which they are frequently in contact with the target language which may allow them to further develop it and provide more tools to the teacher to assess language skills. On the other hand, language assessment and assessment literacy began to develop in ESL contexts therefore teachers in this context may have a different need for more specialized practice of

assessment. In an EFL setting, students are exposed to the target language only during class which may have a direct impact on the nature of the assessment carried out in the classroom. On the other hand, teachers may be required to teach and assess the four language skills which limits class time and time dedicated to assessment. The involvement of teachers and students in institutional assessment decisions is very frequently kept to a minimum, therefore the opportunities for reflective assessment and improvement of assessment procedures may also be kept to a minimum.

Although, students and teachers in EFL/ESL contexts face very similar realities (assessors use rating tools for a variety of purposes, teachers need to be informed of assessment and testing principles to link high-stakes test and the classroom, teachers may be responsible for the production, rating and interpretation of tests, among others) their differences seem to be withstanding. In terms of language assessment literacy, the needs of teachers in both contexts may also converge and diverge greatly. For instance, in both settings, teachers may need support to find ways of contextualizing and situating assessment in their students´ specific context. They may also need to find ways of connecting classroom assessment with high-stakes language testing so students can suit context-specific language policies (Froetscher, 2017). Having said this, it can be safe to state that ESL and EFL assessment have similarities that tie the literature together but major differences that characterize their traits. Thus, it seems relevant to construct assessment literacy literature emphasizing the differences of assessing each language skill since in EFL contexts teachers assess diverse skills in their classrooms. Therefore, this project seeks to fulfil this need of language assessment literature in EFL contexts by seeking to fulfil the research purposes described below.

3.8 Research Purpose

The overall aim of the study is to analyse the impact that two sessions of writing assessment training had on EFL Mexican university teachers. It focused on three main areas of potential impact, a) teachers´ reported classroom assessment of students´ writing skills, b) teachers´, language program managers´ and students´ perceptions towards writing assessment as well as assessment training and c) the changes that training may have encouraged in teachers´ analytic and holistic scoring. These goals were led by five research questions,

1) *To what extent does writing assessment training impact EFL teachers´ reported classroom assessment of students´ writing skills?(RQ1)*
2) *What is the impact of assessment training on teachers´ perceptions of writing and on their perceptions of classroom writing assessment? (RQ2)*
3) *What is the impact of assessment training on language program managers´ perceptions of writing assessment? (RQ3)*
4) *What are students' perceptions of EFL teachers´ regular classroom writing assessment and of the importance of writing assessment training?(RQ4)*
5) *To what extent does writing assessment training and teachers' personal background impact their use of analytic and holistic scoring tools to assess opinion essays in the EFL classroom?(RQ5)*

The overall and specific objectives of this study intend to contribute to the literature in language assessment literacy by emphasizing the need to focalise language assessment literacy on specific language skills rather than a generalized perspective of language assessment as has been approached in previous research. Specifically, this study attempts to address five key gaps in the literature such as a) the lack of analysis of the impact of

assessment training as a contextual factor that may influence EFL classroom assessment of writing (RQ1), b) the insufficiency of research that analyses not only the views of teachers but also the views of language program decision makers and EFL students (RQ2, RQ3 and RQ4), c) the shortage of quantitative perspectives that may give an insight of the role that teacher personal traits may have in the score given to classroom assessment of writing (RQ5) and d) the absence of studies that address EFL assessment and assessment literacy in Mexico.

Studies such as those conducted by Cumming (2001), Cheng *et al.* (2004), Chen *et al.* (2013), Leung and Mohan (2004), Lee (2007), Vogt and Tsagari (2014), Inbar-Lourie and Donitsa- Schmidt (2009) have focused on classroom assessment and how teachers conduct it in an EFL context. Studies such as Chen *et al.* (2013) and Yan, Fan and Zhang (2017) agreed that assessment was deeply influenced by the context in which assessment is involved (economic, social, and cultural factors of the institution). Therefore, it can be argued that research in regard to assessment in FL classrooms has largely focused on how teachers assess their students in an exploratory sense, to understand the nature and purpose of the score provided. However, the specific influence that training, an additional contextual factor, may have on assessment has remained underexplored.

Research has yet to clarify the level of impact that training produces in instructors´ assessment practice, particularly in the EFL classroom in regard to writing performance, to begin to understand the potential value of assessment literacy. It is my belief that the Latin American context has also remained underexplored considering that studies that focus on assessment training and its influence have been conducted in Asia (Koh *et al.*,2017), North America (Nier *et al.*, 2013; Malone, 2013) and Australia (Knoch, 2011). Therefore, this

research project seeks to examine the level of impact that assessment training has on teachers´ classroom writing assessment practice in the Mexican university EFL context by providing answer to the first research question (RQ1) *To what extent does writing assessment training impact EFL teachers´ reported classroom assessment of students´ writing skills?*

In regard to perceptions, research has attempted to understand the perceived levels of assessment literacy of teachers in L1 educational contexts (Metler, 2003; Metler and Campbell, 2005) to further explore the direction of their classroom assessment practices. Other experts have focused on what teachers and language testing experts consider are their assessment literacy needs (Hasselgreen, Carlsen and Helness, 2004, Vogt and Tsagari, 2014) to create courses and/or materials that enhance assessment literacy (Fulcher, 2012). Perception-focused research has also tried to understand teachers´ views of existing assessment training workshops or courses (Nier *et al.* 2013; Malone, 2013; Jeong, 2013, González and Vega López, 2018) viewing them as useful but pointing out that they needed to be complemented with additional practice. It has also been commonly found that language teachers considered they did not have the necessary training to assess their students objectively (López Mendoza and Bernal Arandia, 2009; Volante and Fazio, 2007) or that assessment literacy courses provided to pre-service and in-service teachers were not enough to equip them with the necessary tools to assess their students (Lam, 2015).

It seems that the focus of research has largely been teachers and their needs/views regarding assessment literacy. It is my belief that research still needs to consider other involved stakeholders such as language program decision-makers and students and their opinions of the importance of assessment literacy. Considering their points of view may

allow to have a more complete construct in terms of stakeholders´ needs, factors to consider when assessing language skills, the potential washback of assessment and a detailed plan of how to enhance assessment literacy in teachers. Considering this, the second purpose of the study is to analyse how participants´ views of EFL writing assessment changed post to experiencing training in comparison to their pre-training points of view. Three research questions guide the fulfilment of this purpose, a) *What is the impact of assessment training on teachers´ perceptions of writing and on their perceptions of classroom writing assessment?* (RQ2), b) *What is the impact of assessment training on language program managers´ perceptions of writing assessment?* (RQ3) and c) *What are students' perceptions of EFL teachers´ regular classroom writing assessment and of the importance of writing assessment training?* (RQ4).

A large number of studies that examine the impact of training on assessment, focus on statistical analysis of scores that raters provided prior and/or post to training in L1, L2 or FL large-scale testing contexts (Baily and Brown, 1996; Elder *et.al*, 2005, 2007; Weigle, 1994, 1998, 2007; Shohamy *et al.*,1992; González and Urias, 2009). In FL settings, such as those found in Mexico, instructors judge their students´ performance with a numerical score that later is reported to the educational institution´s administration (Ketabi and Ketabi, 2014) thus acknowledging the importance of understanding scores provided to students´ work. Additionally, studies such as Vogt and Tsagari (2014) ain addition to Volante and Fazio (2007) suggest that teachers do not receive enough training to assess their students thus considering teachers learn about assessment in their daily practice. It is my belief that classroom assessment of writing and how training can be of use to teachers has remained under explored since it is in this setting that not only numerical scores are provided but also feedback to the written task. It can be stated that more research into how

training can actually improve the score and its reliability is needed so that a quantitative perspective of the importance of training for teachers in classroom contexts is provided.

On the other hand, teachers carry their own personal characteristics into the classroom which may also influence scores. Studies such as Barkaoui (2011) and Weigle (1994, 1998) analyse how experienced and non-experienced teachers score writing and the differences found among them. However, it seems that teacher traits such as academic background and gender still need to be explored to analyse if they have a role in FL writing assessment. Therefore, this project seeks to analyse the impact portrayed on the post-training analytic and holistic scores that forty-eight Mexican EFL teachers provided to five sample opinion essays taking into consideration their academic background, gender and teaching experience. The fifth research question (RQ5) *To what extent does writing assessment training and teachers' personal background impact their use of analytic and holistic scoring tools to assess opinion essays in the EFL classroom?* seeks to fulfil this purpose.

From a holistic point of view, this project seeks to contribute to the field of writing assessment literacy by providing an impact categorization of assessment training intervention on the assessment of writing in the EFL classroom. Additionally, it intends to contribute to the field by emphasizing the importance of considering assessment literacy as an area that needs to focus on each language skill to construct literacy of language assessment as a whole. This project contributes to this perspective by providing findings in the area of writing assessment.

From a particular point of view, it seeks to raise awareness, through assessment training in Mexican EFL university teachers in the north-eastern part of Mexico, of the importance of assessing writing in language classrooms to develop students´ language skills. It is considered that this study can provide language teachers and language program managers with the necessary information about teachers´ assessment literacy and its possible outcomes in teachers´ assessment practice. This project also intends to encourage stakeholders to visualize the importance of writing in an EFL curriculum and the need for EFL assessment literate teachers.

3.8.1 Key Theoretical Considerations

As suggested by the updated Model of Writing (Hayes, 2012) portrayed in Figure 3 (p.17) of this thesis, writing is considered a multi-level process in which diverse elements are part of the creation of a written text. These levels include, the Control Level in which the writing plan or scheme is articulated; the Process Level where the writer may embrace distinct roles (translator, evaluator, transcriber, proposer) or where he/she may consider distinct environmental issues such as the text written so far, critics of the text, available materials, and the Resource Level where the distinct working memories are active. However, in reality this process may not always be followed in an L2 or FL classroom. Other contextual aspects of the environment such as lack of teaching time, the students´ learning process, teachers´ teaching approach and the students´ and teachers´ assessment purposes may also have a role in the development of a text therefore suggesting that more levels or a different model should be suggested for the development of a L2 or FL text. Writing is also considered by some experts a social process (White, 1990) in which context, environment and stakeholders are active participants of the final written piece. Therefore, its assessment needs to carry context specific characteristics that link classroom

teaching and learning, with classroom writing assessment and large-scale testing (Froetscher, 2017). Thus, it can be said that the assessment of writing should be included in the writing models since it may have an impact on the production of a text. Research in the area of writing assessment has seemed to largely focus on its subjectivity in large-scale L1 or L2 proficiency tests (Baily and Brown, 1996; Elder *et.al*, 2005, 2007; Weigle, 1994, 1998, 2007; Shohamy *et al.*,1992; Gonzalez and Urias, 2009) without considering what teachers teach and how they assess in the classroom. Therefore, research on classroom assessment of writing may provide the necessary knowledge to consider its inclusion in the previous writing models.

This project accounts for teacher assessment training as one of the main aspects that may impact classroom assessment of writing. It considers the theoretical support of assessment literacy literature (Black and William, 1998a, 1998b; Coombe *et.al.*, 2012; Fulcher, 2012; Inbar-Lourie, 2008; Metler and Campbell, 2005; Stiggins,1995) in which, from a general perspective, it is pointed out that assessment literacy encompasses what teachers know, how they use the knowledge (skills), and their interpretation of their assessment in a specific environment with specific contextual factors. These factors involve teachers´ assessment environment, their assessment perspectives, their students´ needs, students´ views and the assessment policies of the institutions teacher participants are at service in. Nevertheless, this project seeks to contribute to this field by emphasizing the need to consider focalized assessment literacy of each language skill. For instance, writing assessment literacy, speaking assessment literacy, among others.

The theoretical construct of Xu and Brown (2016) is also taken into account as a basis for this project (Figure 6, p. 68) by considering that assessment literacy not only accounts for

the knowledge that teachers may have of assessment and their assessment practices but also with the education they receive and their own reflection processes. Thus, the two training sessions delivered to participating teachers sought to provide them with knowledge, skills, practice, decision-making techniques and opportunities for their own reflection in addition to the previous professional background each teacher may have. The two training sessions, were considered the backbone of the project and a means to collect the necessary data. This data collection procedure is outlined in the following Chapter.

Chapter 4: Methodology

This section presents the procedures followed to collect and analyse data for this project. It begins by describing the methodological approach followed and then goes on to explain the research questions based on the purposes they intend to fulfil. Then, the description of the participants and the context in which they are involved is outlined. The chapter goes on to report the instrumentation used to collect data and the procedures followed to examine it. Finally, specific ethical considerations taken during the development of this study are pointed out

4.1 Methodological Approach

This research project mainly focuses on changes in 1) teachers´ reported writing assessment practices in their EFL classroom, 2) perceptions of stakeholders of the writing assessment process, and 3) teachers´ use of scoring tools to provide scores to student writing.

This study is qualitatively dominant which, from an interpretative constructivist perspective (Johnson *et al*., 2007; Creswell, 2015), seeks to provide my interpretations (as the main researcher) of the realities observed and obtain a more knowledgeable comprehension of the phenomenon under analysis. I seek to go about following this methodological stance by considering that my observations will provide a better understanding when supported by quantitative data (Johnson *et al.*, 2007; Creswell, 2015). Having stated this, it is my belief that a mixed methods approach enables the collection of qualitative and quantitative data in a single study and at different stages allowing for

different scientific inquiry to be approached (Glowka, 2011). Specifically, a convergent design was followed (Creswell, 2015; Teddlie and Tashakkori, 2006) since quantitative and qualitative data were collected separately and at different periods of time throughout the study. Each piece of data was collected and analysed following specific qualitative and quantitative procedures (Perry, 2011) with the purpose of obtaining different perspectives of the situation under analysis and considering that a combination of both approaches may provide a better understanding of a research phenomenon than either approach alone (Creswell and Plano Clark, 2011; Creswell, 2013).

For this project, I considered the research questions of the study as the basis that drove my preference for the use of qualitative (RQ1, RQ2, RQ3 and RQ4) and quantitative (RQ5) approaches to collect and analyse data (Cohen *et al.*, 2011; Onwuegbuzie and Leech, 2005). This methodological stance stems from the interest in triangulating information obtained from both methods. In other words, `...convergence and corroboration of results from different methods studying the same phenomenon´ was sought (Onwuegbuzie and Leech, 2005, p.384) as well as an understanding of the specific object under analysis (Johnson and Onwuegbuzie, 2004).

Specifically, the phenomenon under study is considered to be EFL classroom writing assessment in the Mexican university context. I attempt to study this phenomenon from a qualitative perspective by comparing teachers´ reported assessment of writing prior and post to two training sessions to identify any type of change in their assessment procedures. I also took into account, qualitatively driven, stakeholders´ views of assessment and the impact of training on these. From a quantitative perspective and with the intention of

corroborating previously obtained qualitative results, the scores provided to students´ opinion essays prior and post to training were also analysed (RQ5) considering distinct variables that may have an active role in language assessment such as teacher academic background,

As portrayed on Table 1, multiple data collection instruments were used to collect data that support the mixed-methods pragmatic stance adopted for this project. Data was collected through background questionnaires, pre- and post-training face-to-face semi-structured interviews to EFL teachers and EFL language program managers, an online post-training questionnaire, and student pre-and post-training focus groups.

Qualitative data analysis of transcripts obtained from the semi-structured interviews to teachers and language managers as well as the background questionnaires that participants answered was carried out to answer research question one (RQ1 *To what extent does writing assessment training impact EFL teachers´ reported classroom assessment of students' writing skills?),* research question two (RQ2 *What is the impact of assessment training on teachers´ perceptions of writing and on their perceptions of classroom writing assessment?)* and research question three (RQ3 *What is the impact of assessment training on language program managers´ perceptions of writing assessment?)*. To answer research question 4 (RQ4 *What are students' perceptions of EFL teachers´ regular classroom writing assessment and of the importance of writing assessment training?*) analysis of the transcripts obtained from student focus groups were conducted. Analysis of these multiple data sources allowed the identification of participants' perceptions of writing, writing assessment and assessment training during the 12 months the data collection process

lasted. Additionally, interviews with program managers allowed identifying if by providing assessment training to EFL teachers the teaching and assessment of writing in the Mexican EFL classroom is promoted. Analysis of qualitative data followed a grounded theory approach considering that an `emergent fit´ to data best suited the analysis (Taber, 2000, p.470). In other words, categories and themes that emerged were modified to fit data rather than data chosen to fit each previously stated category. More of this analysis is explained in section 4.8.1 of this thesis.

Differential and inferential data analysis of the analytic and holistic scores provided to five opinion essay samples prior and post to training was used to answer research question five (RQ5 *To what extent does writing assessment training and teachers' personal background impact their use of analytic and holistic scoring tools to assess opinion essays in the EFL classroom?).* This analysis had the purpose of providing a sense of the changes that assessment training may have encouraged in teachers´ scores as well as determine if higher levels of reliability after assessment training are possible. RQ5 also intended to analyse how teacher variables such as gender, academic background and personal background may or may not have influenced their assessment activities.

With the purpose of tracking, examining and interpreting change over a period of time in teacher participants´ assessment practices, I adopted a longitudinal methodological approach (Cohen *et al.*, 2011; Dörnyei, 2007) in which an attempt is made to detect patterns of change and/or explain relationships among variables. The study focuses on the same group of teachers, students and EFL language managers over a period of twelve months and examines the changes that two sessions of assessment training caused in the

writing assessment practices of these stakeholders. Therefore, a panel longitudinal perspective is adopted in which prospective analysis is used to allow (Cohen *et al.,* 2011) examination of ongoing and developing information. Data from a specific group of participants (panel) was collected over an extended period of time. Each time data was collected the same individuals in the panel were analysed. A prospective (Ibid, 2011) type of longitudinal study is considered for this study since it followed and ongoing event and collected information about all the individuals involved in the event analysed (training sessions) as they progressed.

Keeping the main purpose of this study in mind and the epistemological stances I have here expressed, the section below focuses on describing the participants, the data collection instruments, the data collection process and the data analysis procedures.

Table 1 Research Methodology Outline

Research questions	Purpose	Instruments	Analysis Procedures
RQ1 To what extent does writing assessment training impact EFL teachers´ reported classroom assessment of students' writing skills?	To analyse the types of changes that assessment training produces in teachers´ reported classroom writing assessment practice. Writing assessment practices are examined before training and after training is provided.	a) Pre- training face-to-face semi structured interviews to teachers b) Post- training face-to-face semi-structured interviews to teachers.	a) Transcription of interviews; b) Classification and codification of interview transcripts; c) Identification of emerging themes; d) Codification of themes; e) Interpretation of themes and subthemes; d) comparison of emerging themes.
RQ2 What is the impact of assessment training on teachers´ perceptions of writing and on their perceptions of classroom writing assessment?	To identify and compare how teachers perceive the assessment of writing and writing assessment training prior and post to training. To examine if by providing assessment training to EFL teachers the teaching and assessment of writing in the Mexican EFL classroom is promoted.	a) Background questionnaire; b) Pre- training face-to-face semi-structured interviews to teachers; c) Post- training face-to-face semi-structured interviews to teachers d) On-line post training questionnaire	a) Descriptive statistical analysis of closed answers to background questionnaire; b) Identification of patterns among answers; c) Identification of emerging themes among interview transcripts and open questions; d) Codification of information; e) Interpretation of themes and codes found; f) comparison of themes and codes found.
RQ3 What is the impact of assessment training on language program managers´ perceptions of writing assessment?	To examine how language program managers, perceive the assessment of writing and writing assessment training prior and post to training. To identify if assessment training to EFL teachers, promotes the teaching and assessment of EFL writing.	a) Background questionnaire; b) Pre- training face-to-face semi-structured interviews to managers; c) Post- training face-to-face semi-structured interviews to managers.	a) Analysis of open and closed answers to background questionnaire; b) Identification of repeated themes among answers to background questionnaire; c) Transcription of recorded interviews; d) Codification of information; e) Interpretation of themes and codes found; g) comparison of data found.
RQ4 What are students' perceptions of EFL teachers´ regular classroom writing assessment and of the importance of writing assessment training?	To analyse students´ views regarding the assessment of writing, teachers´ writing assessment practices and writing assessment training prior and post to training.	a) Student focus group sessions.	a) Transcription of audio recordings; b) Classification and codification of transcripts; c) Identification of emerging themes; d) Codification of themes and subthemes; e) Interpretation of themes and subthemes; f) comparison of themes.
RQ5 To what extent does writing assessment training and teachers' personal background impact their use of analytic and holistic scoring tools to assess opinion essays in the EFL classroom?	To examine how teacher participants, use and perceive scoring rubrics in their regular classroom writing assessment. To analyse how teachers´ gender, academic and personal background influence analytic and holistic assessment.	a) Writing samples scored holistically and analytically before and after training sessions.	a) Descriptive Statistics; b) Reliability calculations with scores; c) Paired Sample t-test; d) Independent Sample t-test

4.2 The Research Context

This study focused on EFL university teachers who were in service at three public universities or one language institute in the north-eastern region of Mexico. Institution A, Institution B and Institution C are public universities while Institution D a language institute. Institutions A and B are those that have been in operation for the most time (late 1920s the latter and late 50s the former) while Institution C (2006) is a much younger university. Institution B has a multidisciplinary focus providing graduate and undergraduate programs in many different areas such as humanities, social sciences, administration, engineering and health sciences. While C focuses mostly on engineering, B focuses only on professionalizing teachers of elementary schools. Institution D does not provide undergraduate programs; it only provides language lessons to the university students and other adult and young adult students.

In terms of their language programs and language assessment policies, each university follows their own teaching program, teaching methods and assessment criteria. Institution B and C require their undergraduate students to approve all their English or other foreign language courses and to provide official proof of their English proficiency to obtain their degree diploma. A minimum of 450 points on the TOEFL Institutional Testing Program (ITP) is needed to fulfil this requirement. The TOEFL ITP is a series of English language proficiency assessment instruments that allow institutions to conduct affordable, standardized and reliable assessment of non-natives´ English language skills (Educational Testing Service, 2017). It is a high-stakes test administered by the Educational Testing Services that assesses listening, reading and structure and written expression abilities in test takers. It provides the institution the advantage of administering the test in their

facilities with their staff and their own resources (Ibid, 2017) which is the main reason it is widely used in the Mexican context as a tool to prove English language proficiency skills.

Institution A does not require a language test from students to prove their English proficiency but they do need to approve all their English courses. The three institutions require teachers to hand in a score of students´ language performance on a monthly basis. Teachers among these three institutions assess their students following distinct procedures and assessment criteria thus, following unstandardized procedures. Institution D is a language institute, that offers only foreign language programs of English, French and German to university students and external students who wish to take the language lessons. The English program is divided in ten levels beginning at Introductory level and finishing at High Intermediate (From A1 to B2 according to the Common European Framework of Reference) level. Once students finalize Level X, they are given the option of taking the preparation course for the TOEFL Internet Based Test (IBT) or the First Certificate in English (FCE) administered by Cambridge English Language Assessment. Teachers working at this institute are required to hand in a score that includes the assessment of students´ language proficiency of the four skills. The Language Program Coordination provides the teachers with the writing and speaking prompts to assess performance skills and a specific test to use to assess students´ vocabulary, grammar, reading and listening abilities.

Participants of this study were all teaching staff at one of these previously described institutions and agreed in written form to take part in one or all of the stages of this study. A convenience sampling method was used (Dörnyei, 2007) to select participants which

considered participants who were available and willing to take part in the study. The only trait considered for their participation in this study was that they were EFL teachers in service at the moment of the study. Regarding, student participants a snowball effect method was considered (Ibid) which allows for one participant to lead us to the next participant. In this case, teachers lead the researcher to the student participants. Teachers, language managers and student participants are described in the following sections.

4.3 The EFL Teachers

The first group of participants includes 48 EFL university teachers that were in service at the participating universities and the language institute. The majority of the participants (29) were teaching at the language institute while nineteen worked in public universities. These participants were chosen mainly because they were all active teachers teaching EFL to young adults and adult students. They all had in common their interest to improve their assessment practice in their classrooms, their target students, and the type of program they worked with: an English program that demanded teaching and assessing writing skills. All the institutions had more teaching staff working as English teachers and all of them were invited to participate. However, not all of the teachers were willing to be part of the study. Different reasons, such as lack of time, work overload or lack of payment to participate in the project, were stated as reasons to not be part of the study. Eleven of the forty-eight participants agreed to take part in the qualitative longitudinal part of the study (Data Collection Phases 4 and 5). Once again, all the teachers were invited to take part in this second part of the study. However, not all of them agreed to do so. Only those that had the availability and volunteered to participate were considered for the qualitative longitudinal phase of the study.

Participants of Phases One, Two and Three included the forty-eight EFL university teachers whose ages ranged from twenty to fifty-two years. As shown in Table 2, age twenty-two was the most frequent among participants and fifty-two the least frequent: eight teachers in total were twenty-two and one instructor was fifty-two. Most of the participants were females (thirty-one in total), while seventeen were males. In terms of their academic preparation the majority of the participants (eighteen teachers) were BA students who were working as English teachers while pursuing their undergraduate studies, and a minority (fourteen participants) had an undergraduate or postgraduate degree combined with a teaching certification such as the Teaching Knowledge Test or the In-Service Certificate of Language Teaching provided by Cambridge English Language Assessment. Finally, sixteen participants held an undergraduate or postgraduate degree without a teaching certification. Regarding their teaching experience, it was found that thirty-two teachers had been teaching EFL for five years or less, twelve teachers for five to nine years, and finally four teachers were the most experienced with ten to twenty years of teaching experience. This information que be further portrayed in Table 2.

In relation to their teaching of writing and writing assessment practices, seventeen teacher participants (TPs) stated to `always´ assess writing, while `never´ was chosen by two TPs. Similar frequencies were found to instructors´ use of rubrics. Eleven TPs stated to `always´ use rubrics as assessment tools while seven chose the option `never´. Finally, regarding training background, the majority of the participants stated to have no previous assessment training. Twenty-two TPs stated to have had previous training while twenty-one had previous scoring rubric use training. Table 2 depicts the information here described.

Table 2 Teacher Participant Background

TP	Gender	Age	Months TE	Academic Preparation	Institution of Work	Teach/Assess Writing	Use Rubrics	Assessment Training	Rubric Use Training
12	M	31	60	Eng TKT/ICELT	Public Univ.	Always	Always	Yes	Yes
34	M	38	96	BA TKT/ICELT	Public Univ.	Often	Often	Yes	Yes
13	F	26	72	BA	Lang. Inst.	Always	Often	No	No
14	F	36	204	BA TKT/ICELT	Lang. Inst.	Always	Often	Yes	Yes
22	M	28	84	BA TKT/ICELT	Public Univ.	Always	Always	Yes	Yes
20	F	24	12	BA	Lang. Inst.	Often	Sometimes	No	No
5	M	48	96	Eng TKT/ICELT	Lang. Inst.	Always	Often	Yes	No
73	F	26	96	BA	Public Univ.	Often	Often	Yes	Yes
16	M	41	84	BA TKT/ICELT	Lang. Inst.	Often	Always	Yes	Yes
9	M	28	96	BA	Public Univ.	Often	Sometimes	No	No
4	M	29	12	MA TKT/ICELT	Lang. Inst.	Always	Always	No	No
8	F	25	72	BA	Lang. Inst.	Often	Rarely	No	Yes
40	M	21	12	BA Student	Lang. Inst.	Often	Often	No	No
26	M	42	60	BA	Lang. Inst.	Sometimes	Always	No	No
64	F	21	36	BA Student	Public Univ.	Always	Always	Yes	No
319	F	20	5	BA Student	Public Univ.	Never	Never	No	No
307	M	23	18	BA Student	Lang. Inst.	Often	Sometimes	Yes	Yes
306	F	20	18	BA Student	Lang. Inst.	Never	Never	No	No
315	F	24	4	BA	Lang. Inst.	Often	Rarely	No	No
317	M	24	1	BA Student	Lang. Inst.	Often	Rarely	No	No
301	F	21	2	BA Student	Lang. Inst.	Often	Never	Yes	No
303	F	21	2	BA Student	Lang. Inst.	Sometimes	Rarely	Yes	No
305	M	20	6	BA Student	Lang. Inst.	Sometimes	Hardly Ever	No	Yes
318	M	22	36	BA Student	Lang. Inst.	Often	Rarely	No	No
312	M	22	5	BA Student	Public Univ.	Rarely	Hardly Ever	No	Yes
52	F	28	96	MA	Public Univ.	Sometimes	Hardly Ever	No	No
310	F	22	12	BA Student	Lang. Inst.	Often	Sometimes	No	Yes
302	F	23	3	BA Student	Lang. Inst.	Sometimes	Rarely	Yes	No
311	F	22	12	BA Student	Lang. Inst.	Often	Hardly Ever	Yes	No
304	F	21	2	BA Student	Lang. Inst.	Rarely	Never	No	No
303	F	21	2	BA Student	Lang. Inst.	Sometimes	Rarely	No	No
309	F	22	48	BA TKT/ICELT	Lang. Inst.	Always	Always	Yes	Yes
32	M	40	108	MA	Public Univ.	Sometimes	Never	No	No
62	F	26	48	MA	Public Univ.	Sometimes	Always	Yes	No
54	F	25	42	BA	Public Univ.	Always	Often	Yes	Yes
314	F	22	18	BA TKT/ICELT	Public Univ.	Often	Never	No	No
316	F	22	48	BA TKT/ICELT	Lang. Inst.	Always	Always	Yes	Yes
48	F	32	54	MA	Public Univ.	Rarely	Rarely	No	No
325	F	52	240	MA	Public Univ.	Often	Sometimes	No	Yes
37	F	35	144	BA TKT/ICELT	Public Univ.	Sometimes	Hardly Ever	Yes	Yes
23	F	44	120	BA TKT/ICELT	Public Univ.	Often	Sometimes	No	No
42	F	22	12	BA Student	Lang. Inst.	Always	Always	No	Yes
27	M	39	24	MA	Lang. Inst.	Always	Sometimes	No	No
28	F	33	12	BA	Lang. Inst.	Hardly Ever	Never	No	No
68	F	27	6	BA	Public Univ.	Always	Always	Yes	Yes
322	F	36	180	MA TKT/ICELT	Lang. Inst.	Always	Often	Yes	Yes
7	F	40	84	BA TKT/ICELT	Public Univ.	Always	Often	Yes	Yes
313	M	23	24	BA Student	Lang. Inst.	Always	Often	Yes	Yes

Lang. Inst= Language Institute Public Univ= Public University TE= Teaching Experience

As depicted in Table 3, eleven of the forty-eight participants continued participating in Phases Four and Five, by being interviewed once more after the second training session. Four were males with their ages ranging between twenty-four and forty-five years old. The least experienced male was TP313 who had two years of teaching experience while the most experienced was TP32 with nine years of teaching experience. Two male and three

female participants had a BA combined with a teaching certification (ICELT), two females and one male had an MA and one male was a BA student. Two of the males were in service in a public university and the remaining two in the language institute. Nine of the thirteen participants were females whose ages ranged from twenty-two to fifty-two years of age. The most experienced teacher was TP31 who had been teaching English for more than twenty years, while the least experienced was TP315 who had been teaching for less than a year.

Table 3 Teacher Participants Phases 4 and 5

TP	Gender	Age	Months TE	Academic Preparation	Institution of Work	Teach/Assess Writing?	Use Rubrics?	Assessment training	Rubric Use Training
22	M	28	84	BA TKT/ICELT	PU	Always	Always	Yes	Yes
73	F	26	96	BA	PU	Often	Often	Yes	Yes
16	M	41	84	BA TKT/ICELT	LI	Often	Always	Yes	Yes
315	F	24	4	BA	PU	Often	Rarely	No	No
32	M	40	108	MA	PU	Sometimes	Never	No	No
62	F	26	48	MA	PU	Sometimes	Always	Yes	No
316	F	22	48	BA TKT/ICELT	LI	Always	Always	Yes	Yes
325	F	52	240	MA	PU	Often	Sometimes	No	Yes
37	F	35	144	BA TKT/ICELT	PU	Sometimes	Hardly Ever	Yes	Yes
23	F	44	120	BA TKT/ICELT	PU	Often	Sometimes	No	No
313	M	23	24	BA Student	LI	Always	Often	Yes	Yes

LI= Language Institute PU= Public University TE= Teaching Experience

All of the participants claimed to teach writing with different degrees of frequency. Regarding assessment training, four participants stated to have never experienced training to assess a text while the rest claimed to have minimal experience with assessment training.

4.4 EFL Program Managers

The second group of participants, as outlined in Table 4, was the language coordinators of the EFL programs (PM) in which participants were teaching. Four PMs participated in the study by taking part in two face-to-face semi-structured interviews prior and post to the training sessions. These participants did not participate in the training sessions nor did they score the written samples. Three were experienced, female EFL teachers who had been teaching English for more than 20 years while one (also a female) had been teaching English for less than 10 years.

Table 4 Language Program Managers

LPM	Gender	Age Range	Months TE	Academic Preparation	Institution of Work
1	F	51-70	5	BA	Public Uni.
2	F	51-70	23	MA COTE	Public Uni.
3	F	51-70	28	BA	Lang. Inst.
4	F	30-50	56	MA	Public Uni.

All PMs were females and fulfilled teaching practices while administering the program. PM1 was the least experienced, with less than ten years teaching while the rest had more than 20 years of teaching experience. PM1, PM3 and PM4 were heads of the language programs in the public universities while PM2 in the language institute. PM1, PM3 and PM4 had a master's degree in education but language teaching while PM2 had a BA in English language with a teaching certification (Certificate of Overseas Teachers of English) granted by Cambridge English Language Assessment.

4.5 EFL Students

The third group of participants included five to ten EFL students who took class with one of the interviewed participating teachers. These students participated in two student-centred focus group sessions conducted prior and post to teacher assessment training. The learners were enrolled in the English courses that the participating instructors were teaching at the time of the study. For each of the eleven interviewed teachers that took part in all the stages of the study, one student focus group was considered. Therefore, in total eleven groups were interviewed twice (once prior to training and once after training). The researcher requested the participating teachers´ authorization to approach students and invite them to take part in the study. A segmentation strategy (Dörnyei, 2007) was used to recruit volunteer students with similar age, similar EFL learning experience, and willingness to provide their insight for the project. In other words, this strategy was used to obtain group homogeneity and intergroup heterogeneity (Dörnyei, 2007, p.145).

These Mexican students were of beginner to high intermediate English proficiency level and their ages ranged from eighteen to thirty years old. They were all willing to provide their insight in relation to writing assessment and their teachers' assessment practices in the classroom. They also had in common the program they were studying at the moment and they all were adults or young adults interested in improving their use of English. By considering these homogenous characteristics, I intended to care for the dynamics of the group (Dörnyei, 2007) and to provide better opportunities of obtaining the necessary data to accomplish the purpose of this project.

4.6 Data collection instruments

Quantitative and qualitative approaches to research were followed to collect and analyse the data for this study. This section portrays the instruments used to gather the data of this study in the chronological order in which they were implemented. It first begins by describing the background questionnaire and the pre-training interviews conducted with the eleven teachers and the four language managers. It also describes the student focus groups. It then describes the analytic and holistic rubrics used to score sample papers and the two assessment training sessions provided to participants. Finally, the post-training interviews and the on-line post training questionnaire are outlined.

4.6.1 Background questionnaire

During the first phase of the data collection process, participants were asked to answer a background questionnaire with the purpose of obtaining more information about their experience with the teaching of writing, its assessment and their general EFL teaching background (Cheng, Horwitz and Schallert, 1999; Gardener, Masgoret and Tremblay, 1999; Kitano, 2001). According to Taylor-Powel and Renner (2000), the importance of a background questionnaire is born from the need to understand the target group of participants but above all it helps to understand if the intended audience was reached. Keeping this in mind, two different background questionnaires were used for this study: one for teacher participants´ and program managers (BQ1, Appendix A) and a second one with students who took part in the focus groups (BQ2, Appendix B).

As portrayed in Appendix A, BQ1 included eight closed-ended multiple-choice questions and three open-ended questions. By using an instrument that included both types of

questions, participants had the opportunity of expressing their ideas freely while the data collection and analysis processes were facilitated (Nunan, 1992). Questions elicited participants´ age, educational background, their previous experience with assessment training as well as their initial views of writing and of writing assessment. Prior to its use, BQ1 was piloted (Dörnyei, 2003) with a group of English teachers that were not part of this study with the purpose of obtaining feedback on its use and to determine if its purpose was being fulfilled. Minor changes to format and word order were made to the final draft.

BQ2 (Appendix B) included questions that could lead the researcher to understand students´ EFL learning experiences, experiences with writing assessment and their perceptions of writing assessment. It also explored students´ general background. This questionnaire included open-ended and closed-ended questions (nineteen questions in total) with the purpose of facilitating the analysis process of these responses and obtaining participants rationale to their responses (Taylor-Powel and Renner, 2000; Dörnyei, 2003). As with BQ1, BQ2 was also piloted with a group of students and minor changes to format, word order and typos were made to the final draft.

4.6.2 Interviews to teacher participants

With the overall goal of examining the specific changes that assessment training had on teachers´ writing assessment reported practice, two semi-structured interviews were conducted (one prior to the first training session and a second one after the last training session). An interview protocol (Appendix C) was followed (Creswell, 2013) while being conducted with eleven of the forty-eight teacher participants. The interviews had as secondary goals to a) confirm the context in which EFL teachers worked, b) determine if

teachers assessed writing in their EFL classrooms and their reasons for doing so, c) explore their use and perceptions of the rubrics used for the study, d) analyse their perceptions of the training session provided and finally e) analyse the extent to which the teaching and assessment of writing is promoted through assessment training. The eleven participants volunteer to take part in this phase depending on their availability to be interviewed and their language teaching background. Interviewing participants of distinct backgrounds and teaching experience had the purpose of obtaining a diverse set of data that could provide a complete picture of their regular assessment practices.

Interview 1 (prior to training, Appendix C) intended to obtain a sense of teacher participants´ actual assessment practice prior to the assessment training. It also aimed to understand the context in which the teacher was working. It included thirteen questions in Spanish, teachers´ L1. The use of participants´ L1 combined with the semi-structured format of the interview had the purpose of providing interviewees with a comfortable environment in which the researcher could explore data while providing direction and guidance with an interview outline (Dörnyei, 2007). With the purpose of allowing interviewees feel more comfortable with their natural language (Cohen *et al.*, 2011) and to avoid transcript translation diminish data objectivity (Pavlenko, 2007), the interviews were conducted in the language of the participants´ choice, being English or Spanish the options offered. The researcher is a fluent speaker of both languages therefore analysing information in both languages was feasible. In total three interviews were conducted in English while eight in Spanish. Interviews were audio recorded and then transcribed for further analysis. Interview 1 (Appendix C) was piloted before its use. It lasted from 20-30

minutes and was conducted from two to three weeks prior to the first assessment training session.

Interview 2 (post to training, Appendix D) had the overall goal of discovering the changes, that according to the participants´ perception, the training had encouraged. It focused on the teachers´ regular assessment practice, their use and perceptions of analytic and holistic scoring tools and of writing assessment. Additionally, it intended to obtain data to compare with the information obtained from Interview 1. Interview 2 was conducted post to the second session and towards the end of the study to allow teachers to reflect on the information shared during the sessions and to implement any changes they considered necessary in their regular classroom assessment of writing. It also followed a semi-structured format that guided the researcher during the session allowing for flexibility in the answers provided by participants. As with Interview 1, participants were given the option of choosing the language of their preference being English and Spanish the two options available.

The researcher decided to analyse information obtained from transcripts in the language they were collected considering that research interviews should use interviewee´s natural language to gather and understand qualitative knowledge and to avoid the influence of translation on data bias or subjectivity (Pavlenko, 2007). Similar to Interview 1 (Appendix C) three interviews were conducted in English while eight in Spanish. Interviews were audiotaped and then transcribed for further analysis. The interview outline was piloted twice before it was used with participants. The piloting stage was conducted with one experienced and one inexperienced EFL teacher that were not part of this study with the

purpose of obtaining points of view from people with similar backgrounds to those participating in this study. Changes to the outline after piloting included rephrasing of questions to make them clearer and elimination of questions that elicited repeated information. Each interview lasted from 20-30 minutes.

4.6.3 Interviews to language program managers

Language program coordinators or managers (PM) are the decisions makers of the English programs at the institutions under analysis. Two interviews were conducted with each of the four managers.

Interview 1 (Appendix E) had the intention of exploring the managers´ professional background, their opinions and perceptions of the language programs with which they were working at the time of the study. It also had the intention of exploring the issues that managers experience when including writing as a component of the language program they administer and including its assessment in the EFL curriculum. This interview was conducted before Training Session One to allow the researcher interpret the characteristics of the program and those of the decision makers prior to the training intervention. It followed a semi-structured format including twelve open-ended questions that were previously designed in an interview protocol (Cohen *et al.,*2011) included in Appendix E. It was piloted with an EFL language program manager who was not part of this study. Once piloting finished, order of questions, and word order in questions were improved. Additionally, two questions were added to the interview outline: a) Did the training session provided help improve the management of the program? If so, how? and b) Do you consider the training session helped the teachers of your program improve their everyday

practice? If so, How? Interviews lasted from 20-30 minutes. They were recorded and transcribed for further analysis.

The second interview (Appendix F), was conducted after the second training session with two main purposes: 1) to identify any changes that had occurred post to training in managers´ perceptions of writing assessment and of the importance of providing training to teaching staff; but above all 2) it sought to examine if writing assessment training can raise awareness of the importance of writing assessment in Mexican EFL programs and classrooms. This interview followed the same process for its validation than that followed for Interview 1. After being piloted minor format changes were implemented. Both interviews, were audio recorded with the use of a digital recorder with previous consent of the interviewees for future transcription and analysis.

4.6.4 Student Focus Groups

Focus groups (Appendix G and H) refer to the type of unstructured interview in which a group of people are `…often accompanied by a facilitator whose goal is to keep the group discussion targeted on specific topics...´ (Mackey and Gass, 2005, p.173). A focus group has the purpose of providing a friendly environment so that participants can engage in group brainstorming and react to issues and topics that arise in the discussion (Dörnyei, 2007).

Therefore, for this study the focus group interviews had the purpose of hearing EFL students´ views and perceptions in terms of writing assessment, teacher assessment training and their teachers´ writing assessment in their classroom in an environment in

which students could feel comfortable expressing their ideas. To provide students with comfort and security during the sessions and avoid any discrepancy or data loss (Mackey and Gass, 2005), sessions were conducted in Spanish and audio recorded for further analysis. Sessions followed a semi-structured protocol (Appendix G and Appendix H) that included closed and open-ended questions to obtain a wide range of information and adapt to responses obtained from participants. Every student taking the interviewed teachers´ class was invited to take part in two sessions: the first prior to the assessment training session one and the second post to training session two. Five to ten participants volunteered to participate. The flexibility and interactivity (Mackey and Gass, 2005) of these sessions allowed the researcher to obtain, according to students´ insight, any changes in teacher participants´ classroom assessment practice prior and post to the training sessions.

4.6.5 Writing Assessment Training Sessions

The writing assessment training sessions are considered the core of this study. Two sessions were provided to teacher participants, one at the beginning of the study and the second half way through the project (from six to eight months after session one).

The first session focused on the analysis of general aspects of the nature of EFL writing, its assessment and included a session of writing assessment practice using the holistic and analytic rubrics developed for this study. The second session focused on updating the previous information reviewed in session one. Additionally, it gave priority to the practice and importance of using a rubric as a classroom tool to assess writing and provide formative feedback to students. It also included opportunities for teachers to reflect on

their own context and current assessment processes to analyse how they can be improved including special emphasis on the concept of `assessing for learning´ (Stiggins, 1995).

Prior to the implementation of the training sessions, these were piloted with a group of English teachers part of the staff of the Language Institution part of this project. Thirty-one teachers attended the training session that lasted three hours with a twenty-minute break between the theoretical/discussion part and the scoring practice session. Although this session was considered the piloting of the first training session; the content, practice and the attendees that completed this first phase of the project were considered valid data of this study since minor changes to the management of teachers´ participation were made. For the pilot session, teachers were asked to bring the five writing sample papers scored analytically and holistically. Participants who did not assess their five samples prior to the training sessions (approximately 20 teachers) were not considered in the study.

With the purpose of providing teacher participants with distinct dates and opportunities to attend the training sessions, session one and two were offered on three different occasions. Each session was carried out in a distinct institution and on a distinct date to give more freedom to teachers of choosing which session to attend. Each session lasted two days and approximately two and a half to three hours each day. They were conducted following the same content and structure to avoid having distinct content influence the impact of the sessions. Structure and content of training sessions are described in the following paragraphs.

The structure of sessions 1 and 2 followed the content suggested in the CEFR Manual for Language Examinations (Council of Europe, 2002, 2009a, 2009b), the ALTE Manual for Language Test Development and Examining (Council of Europe, 2011) and the principles suggested by Bachman and Palmer (2010). The manuals suggest that assessors undergo a) guided discussions of samples that are already scored; b) participate in independent marking and follow-up discussions of discrepancies found among scores; c) conduct independent marking and pair discussions of scores given (Council of Europe, 2011).

Although the principles outlined in the manuals are oriented towards large-scale testing settings, the researcher adapted them to the Mexican classroom setting in which teachers need to hand in a score of students´ performance to the institution´s administration. To do this, strategies to incorporate formative feedback in combination with the summative scores were discussed with the participants. From the principles outlined by the Council of Europe, group discussions of scores and independent scoring practice were adapted to the sessions provided considering the teachers´ classroom nature, needs of students and institutional requirements. The steps and suggestions provided by Bachman and Palmer (2010), discussed in section 3.4 of this thesis, were also considered for the elaboration of the training session content in the sense that group discussions of the understanding of analytic and holistic scoring tools, discussion of scores provided to benchmark papers, and paper scoring practice were incorporated to the sessions.

During day one of session one, participants were provided with the written samples (Appendix I) to assess and the analytic (Appendix J) and holistic (Appendix K) rubrics to use for scoring. During this session participants were explained the nature of their

participation and assessed each paper independently without the intervention neither of the researcher nor of any other participant. Day two was divided in three phases. Theoretical background to assessment was discussed during Phase One while during Phase Two benchmark written samples that included a score were analysed. Phase Three encouraged teachers to practice scoring texts. A more detailed description of these phases is included below.

a) Phase 1. A theoretical discussion was encouraged which included a) the differences among evaluation and assessment; b) the importance of evaluation and assessment to a language program and to EFL teachers´ daily teaching practice; c) the difficulties found among classroom assessment; d) the difficulties of assessing writing; e) the difference among distinct types of rubrics and f) the variability and reliability of scoring writing. The researcher/trainer used different techniques such as participant elicitation, group discussion of information, and visual support with technological tools such as PowerPoint. This first phase lasted approximately sixty minutes followed by a fifteen to twenty-minute break.

b) Phase 2. For this phase, the researcher/trainer led participants in an analysis of benchmark written samples which initiated with a presentation and discussion of three scored writing samples. These samples were obtained from the Longman TOEFL preparation Book (Phillips, 2009) with the purpose of having an official score and official explanations to the scores given to each paper. Scores and discrepancies among scores given to each paper were discussed among the attendees. Three more scored samples obtained from foreign EFL intermediate students were also analysed. These samples were scored with the rubrics adapted by the researcher for this project. This second analysis had

the purpose of being able to discuss distinct levels of proficiency of written texts. Then, teacher participants shared their insights on the scores given to each paper. This phase took approximately 30-40 minutes.

c) Phase 3. This last phase focused on participants´ scoring practice of three samples provided to the attendees by the researcher/trainer. Writing samples were not previously scored with the purpose of having teacher participants assess them independently. As in the previous phase, samples were of different proficiencies. In pairs, participants assessed the first two papers by giving a holistic and analytic score based on the rubrics provided and by giving any comments the teacher considered necessary to add to the score. Once finished, scores given by each pair of participants were discussed among all the attendees with the purpose of analysing discrepancies and difficulties that may have arisen while using the rubrics. Discussion was guided by the researcher/trainer and took approximately 40-50 minutes.

Session Two focused mostly on providing teacher participants with some insight and opportunities of reflection in terms of the current role of assessment in their classrooms, as well as sharing their experiences with the assessment of writing. The possibilities of integrating the concept of assessing for learning in their assessment practice were also explored. This intervention took place approximately six to eight months after the first session was given. For approximately two to three hours, the researcher/trainer encouraged teachers to analyse their context and their current assessment practices with the purpose of reflecting on which assessment practices were most suitable for their students and their

assessment purposes. Elicitation techniques and group discussions were encouraged to facilitate self-reflection.

4.6.6 The Written Samples

Each instructor was provided with five opinion essay samples to score pre- and post to the assessment training sessions. The samples were scored with an analytic and holistic rubric (Appendix J and Appendix K) specifically adapted for this study. The five samples were scored on two occasions: once before assessment training session one and a second occasion post to assessment training session two. Samples were written by EFL low intermediate university students enrolled in a Mexican public university. The writing task required students to write their opinion about a specific statement in minimum 120 to maximum 180 words. The task prompt and the written samples may be found on Appendix I.

4.6.7 The scoring rubrics

The main purpose of this study is to analyse how assessment training impacted the classroom assessment of writing of Mexican EFL instructors. It also seeks to understand how EFL teachers´ background has a role in their holistic and analytic assessment with the use of scoring rubrics as assessment tools.

In the Mexican context, rubrics are often used not only in large-scale testing contexts but also in classrooms to guide and focus teachers´ assessment. With the intention of fulfilling this project´s purpose, the teacher participants of this study used two scoring rubrics to provide scores to the five written samples previously described: an analytic and holistic

rubric (Appendix J and Appendix K) specifically adapted for this study. Although there are different types of scoring rubrics that can be used in writing assessment, these two types were chosen since they are the most widely used in the Mexican EFL university and language institute contexts.

The analytic scoring scale included a description for five different scales, being five the highest and zero the lowest score. This analytic scoring tool was adapted by the researcher considering as a main basis the standards set by the Common European Framework (Council of Europe, 2002), the Manual for Language Test Development and Examination by the ALTE and the Council of Europe (2009a, 2009b), Jacob´s *et. al* (1981) as well as Weir (1990) rubrics. Adaptation had the purpose of making the rubric as easy and clear as possible to minimize the issues that inexperienced participant teachers may encounter during the scoring process. Once it was adapted, the scoring scale was piloted with three experienced language teachers. Although their review of the tool was done independently, the three teachers agreed in their suggestions. Initially, the scoring rubric considered six categories to be assessed. However, reviewers agreed that five categories would be enough to assess writing because by including a sixth a risk of category misunderstanding on behalf of inexperienced teacher participants could arise. On the other hand, the researcher considered that inexperienced teacher participants could benefit from a simple and clear analytic rubric while experienced ones could encounter an easier task when using this rubric. A second premise considered when deciding the number of categories and the number of scale levels to describe in the scoring tool was the consistency and the meaningfulness of the scoring scale (Bachman and Palmer, 2010). In other words, it was considered that by including only five categories to assess instead of six the consistency of

scores and the meaningfulness of the descriptors to teacher participants would be accounted for.

The holistic scoring scale contained a five-point scale description that focused on the positive traits of students´ texts. The adaptation of this holistic scale consisted in the same process followed for adaptation of the analytic rubric described in the previous paragraphs. Other holistic rubrics such as those outlined in the Common European Framework (Council of Europe, 2002), the Manual for Language Test Development and Examination by the ALTE and the Council of Europe (2009a, 2009b) and the IBT Next Generation TOEFL Test Independent Writing Rubrics (2004) were used as a baseline for the adaptation process. Once adapted, the scale was shared with three different teachers to obtain feedback that could aid in its improvement. The teachers were given a two-week period of time to use the rubric and analyse its structure. This piloting phase was done simultaneously as with the piloting phase of the analytic rubric and considered the insight of the same teachers. The researcher decided to keep a five-point scale and their corresponding description to balance its use with the analytic rubric and based this choice on the optimistic reliability of `a writing test that is able to distinguish reliably between five scale points or more´ (Weigle, 2002, p.123).

4.6.8 Post-training online questionnaire

Once participants completed both training sessions, the pre-scoring and post-scoring process, the forty-eight initial teacher participants of the study were asked to answer a post-training questionnaire (Appendix L). This questionnaire was delivered in Spanish and had the main goal of eliciting participants` perceptions in relation to the usefulness of

training and the use of scoring tools for classroom assessment purposes. The questionnaire was delivered electronically through an online survey platform (Isurvey https://www.isurvey.soton.ac.uk/) provided by the University of Southampton with the purpose of making the data collection processes more effective for the researcher and attractive for the teacher participants (Dörnyei and Taguchi, 2010).

The link to the survey was sent to each participant via email. Teachers answered the questionnaire from two-three weeks after completing the final scoring (post to training two) round. This questionnaire included a total of fourteen open-ended and closed-ended questions. The first nine questions were closed questions which provided a five-level Likert scale, a group of answers in which participants were asked `to indicate the extent to which they agree or disagree with it (the statement) by marking…one of the responses ranging from strongly agree to strongly disagree´ … (Dörnyei, 2007). The following five items on the questionnaire combined answer choices and opportunities for participants to provide an explanation for their answers. The questionnaire protocol that was uploaded to the Isurvey platform is found on Appendix L.

4.7 Data Collection Procedures

To fulfil the purposes of this study, different procedures were followed to collect the data. This section provides a chronological description of the five stages that were followed to collect data with each of the instruments previously described. Table 5 includes a chronological overview of the procedures followed.

4.7.1 Stage 1 Teacher, Language Manager and Student Interview 1

During the first phase participants were explained the nature of their participation. This explanation varied depending on the role of the participant in the project. While teachers engaged in two training sessions, scored writing samples, participated in two interviews and answered an online questionnaire. EFL program managers also took part in assessment training and were interviewed twice. Students on the other hand, took part in two sessions of focus groups.

Once this explanation was given, those who agreed to participate signed an informed consent (Appendix N), and filled-in their corresponding background questionnaire (Appendices A, B). This process was conducted during Interview 1. The same process was followed with the first interview to the language program managers and the first focus group with students. The researcher individually contacted teachers and managers via email, telephone or Facebook inbox message to schedule each interview and the first assessment training session. Once interviews were concluded, and with the help of the teacher participants, the student participants were contacted to schedule focus group sessions depending on the students' availability. Interviews and student focus groups were audio recorded, with previous consent of the participant, on a digital recorder.

4.7.2 Stage 2 Assessment Training 1 and Scoring Round 1

This stage focused on the initial assessment training session (AT1). Teachers were provided with a folder that included the writing samples, a copy of each rubric and a set of scoring instructions. Teachers were instructed to individually provide analytic scores, holistic scores and written comments to the papers. Scores were recorded directly on the

format provided with each sample paper. Written samples were scored on site during day one of the training session to avoid material being lost or forgotten by participants. Once participants completed the scoring process, each one handed in their folder and were asked to come the next day for the second part of the training session.

On day two, teacher participants attended the training session during which the researcher provided general background to language assessment, benchmark samples to analyse as a whole and opportunities for independent and group scoring.

4.7.3 Stage 3 Assessment Training 2

This third stage focuses on the second assessment training session which took place approximately six to eight months after the first session, the second session of training one was provided. During the session, participants engaged in further assessment practice with the scoring rubrics. For approximately three hours teachers were encouraged to share their reflections in relation to changes they noticed in their assessment of writing since they experienced assessment training one. Participants were encouraged to participate freely, to interact and share experiences in a group-led discussion.

Once this second training session finalized, each participant was provided with the same five written samples to be scored individually. Papers were provided in distinct order as in the pre-training scoring phase to avoid having teachers remember the scores given to each paper in the first round of scoring therefore avoiding score bias. Participants were asked to take the final samples home and to

score them independently. As in the first round of scoring, teachers were instructed to record their analytic and holistic score on the copies provided and also record any additional written comments they would give to the paper.

Participants were given approximately two to three weeks to rescore papers and return them to the researcher. Then, participants were contacted again to collect scored samples and request their availability to be interviewed once more. Teachers who agreed to continue taking part in the study (eleven teachers) were scheduled to be interviewed once more and those who did not wish to continue were not considered in the data collection. During this sample collection meeting, teachers were notified that they were sent a link via email to answer the online post-training questionnaire. Teachers were explained the purpose of this questionnaire and how to access it on the Isurvey platform provided by the University of Southampton. The email sent to each participant explained the structure of the questionnaire, the instructions to answer it, and provided a participant ID number so the participant could be tracked on the Isurvey platform without revealing their identity. The online questionnaire was answered individually by each teacher approximately two weeks after training session two.

4.7.4 Stage 4 Teacher Interview 2 and Scoring Round 2

This fourth stage had the purpose of collecting the samples scored in the second and final round of assessment to obtain an insight on how teacher participants´ considered their classroom assessment of writing had changed after taking the two training sessions.

This stage initiated two weeks after the second assessment training was completed. During this time, participants had the opportunity to score papers and reflect on their EFL classroom assessment practices. The second interview to teachers was conducted two to three months after training two concluded. Teacher participants were contacted individually via email or Facebook inbox message to schedule it. Interviews lasted approximately twenty-five to thirty-five minutes and were conducted by the researcher. Interviewees´ consent was requested for the interview to be audiotaped on a recording device for future analysis. Once this second interview was done, teachers who had their student writing samples ready handed them into the researcher in print. The rest returned them on a different occasion in agreement with the researcher.

4.7.5 Stage 5 Student Focus Group 2 and Language Manager Interview 2

This collection phase intended to, with the student focus groups and manager interviews, obtain data that could provide an insight in relation to the changes in participant perceptions of classroom writing assessment. Secondly, this stage intended to collect data that could confirm or not the usefulness of assessment training to promote writing assessment in Mexican EFL classrooms.

The second student focus group was conducted with the same students that participated in the first group and the same moderator (researcher) to facilitate interaction and flow of information. Students were contacted once again through their participating teachers to agree on a date for the group to get together. Contact was done via email or Facebook inbox. The session followed the Student

Focus Group Protocol 2 (Appendix H) and lasted from thirty to forty minutes. During the session, student participants were elicited the questions on the protocol while answers from all the students were encouraged with the purpose of allowing all the participants provide their points of view. The session was audio recorded for future transcription.

The second interview to language managers was also conducted in this stage. Once again language managers were contacted via email, telephone or Facebook inbox message to schedule a final meeting. The interview followed the same process as the first interview but was guided with a distinct protocol (found in Appendix F) and was concluded in twenty to thirty-five minutes. As stated in section 4.7 of this thesis, Table 5 outlines the chronological sequence of the distinct data collection stages.

Table 5 Data Collection Procedure

Stage	Activity/Instrument	Participants
1 **Apr-May 2015**	Consent Form Background questionnaire Teacher Interview 1	Teachers
	Language Program Manager Interview 1 Student Focus Group 1	Program Managers Students
2 **June-Dec 2015**	Assessment Training 1 Day1: Round 1 of Sample Scoring Day 2: Training Session/Collection of Samples	Teachers
3 Jan- Feb 2016	Assessment Training Session 2	Teachers
4 **Mar-Apr 2016**	Round 2 of Sample Scoring Teacher Interview 2 Teachers Answer Online Questionnaire	Teachers
5 **Apr - Jun 2016**	Language Program Managers Interview 2	Program Managers
	Student Focus Group 2	Students
	Collection of Scored Samples Round 2	Teachers

4.8 Data Analysis Procedures

This project considers the centre of the study the two writing assessment training sessions provided to the participating EFL Mexican teachers. It analyses the impact that these had on EFL teachers´ assessment of writing in three broad areas: a) writing assessment procedures in their EFL classroom, b) teachers´, language managers´ and students´ perceptions of writing assessment, c) teachers´ analytic and holistic scores to writing. Since a mixed-methods stance with predominance on qualitative data is being adopted for this project, the following section gives an account of the qualitative and quantitative procedures followed to analyse the data obtained. It firstly describes the qualitative analysis and continues with the quantitative one.

4.8.1 Qualitative Analysis

To answer RQ1 (*To what extent does writing assessment training impact EFL teachers´ reported classroom assessment of students' writing skills?),* RQ2 (*What is the impact of assessment training on teachers´ perceptions of writing and on their perceptions of classroom writing assessment?),* RQ3 (*What is the impact of assessment training on language program managers´ perceptions of writing assessment?)* and RQ4 (*What are students' perceptions of EFL teachers´ regular classroom writing assessment and of the importance of writing assessment training?)* qualitative strategies were used to analyse the data obtained from the transcripts of teacher and language program manager semi structured interviews and the transcripts of student focus groups. The same process was followed to analyse answers to open-ended questions on the post-training online questionnaire. This analysis followed a grounded theory approach which according to Strauss and Corbin (1994) uses the constant comparison of data obtained from the

collection sources to generate theory. This theory is generated with the researcher´s interpretations of the voice of the people involved in the study to understand their individual actions (p. 274). In other words, its main purpose is to provide an understanding of the underlying reasoning `grounded´ in participants´ rationale to their actions and practice (Lingard, Albert and Levinson, 2008; Taber, 2000). According to Charmaz (2008) grounded theory is considered an emergent method since,

> `it starts with a systematic, inductive approach to collecting data and analysing data to develop theoretical analyses. The method also includes checking emergent categories that emerge from successive levels of analysis through hypothetical and deductive reasoning´ (p.155).

Therefore, since this study did not rely on a set of pre-established themes and categories into which data was fixed but instead established themes and categories as they emerged in data, a grounded theory approach to the analysis of information was considered. In other words, analysis was data driven.

Considering that interviews are a flexible tool that allows the interviewee to provide personal interpretations or points of view of the world and context in which they interact (Cohen *et.al*, 2011), transcripts of interviews were examined following an interpretative approach of analysis and considering the information obtained from the interviews as a holistic narrative of the participants´ view of the phenomenon of EFL writing assessment. They were analysed in the language they were collected considering that research interviews may use the interviewee´s natural language to collect and comprehend qualitative data (Cohen *et. al*, 2011) and to avoid the influence of translation on data bias or subjectivity (Pavlenko, 2007). Therefore, six interview transcripts (corresponding to

three participants) were analysed in English while the rest in Spanish. Once data was transcribed, information was reviewed and emerging themes were identified. Main themes identified were clustered into subthemes and then into categories with the purpose of noting relationships among variables and the context in which participants were immersed. Each category was coded and then frequencies for each code were obtained (Creswell, 2015). Then, emerging themes, subthemes and categories for each participant were compared to those of the rest of the interviewees to find similarities and differences.

4.8.2 Quantitative Analysis

RQ5 *(To what extent does writing assessment training and teachers' personal background impact their use of analytic and holistic scoring tools to assess opinion essays in the EFL classroom?)* seeks to a) examine if participants´ analytic and holistic scores significantly changed post to receiving training b) analyse if reliability levels of the scores improved post to assessment training and finally c) analyse if participants´ personal background information such as gender, teaching experience and academic background influence their analytic and holistic scores.

This quantitative aspect of the study provides a wider perspective of the level of impact that other contextual factors may have on teachers' classroom assessment of writing. Therefore, specific statistical analysis to establish correlations among the scores was run with the aid of the statistical software programme SPSS v.23. Scores given to the five opinion essay papers were introduced to SPSS then calculations were run to obtain means, modes, minimum/maximum scores, standard deviation and frequencies of each score with the purpose of describing and presenting the data (Cohen *et. al*, 2011).

To analyse the reliability of the scores provided prior and post to training, an Intraclass Correlation Coefficient (ICC), specifically a two-way mixed method, was used to calculate the reliability of scores among teacher participants. The ICC performs an analysis in which two or more scores are provided to the same subjective matter (Leech, Barrett and Morgan, 2014) and it was considered that a strong correlation is shown if two or more scorers agree in the scores provided to the same phenomenon (Roever and Phakiti, 2017).

Then, a Paired Sample T-test was calculated to determine the extent to which scores differ prior and post to training and if the differences found were significant (Woodrow, 2014). This calculation was run with the forty-eight participants and their sets of scores prior and post to training (total of ninety-six scores). Finally, with the purpose of understanding the degree to which distinct teacher characteristics had a role or were independent from scores provided (Bachman, 2004), an Independent Sample T-test (Woodrow, 2014) was run. To do so, the forty-eight participants were divided in three groups according to the following characteristics: a) teaching experience, b) academic background and c) gender. Among these three categories, participants were grouped in subcategories according to their unique characteristics: a) males and females (gender); b) teachers with five years or less of teaching experience and teachers with more than five years of teaching experience (teaching experience) and c) teachers who were undergraduate students and those that already had an undergraduate degree at the moment of the study (academic background).

In relation to the answers provided to the background questionnaire, data was revised to determine if answers were complete, accurate and uniform among all respondents (Cohen *et. al*, 2011). Considering that a code is a label provided to a piece of data either decided in

advance or in response to data found (Cohen *et. al*, 2011), possible answers to closed questions were pre-coded (given a code before participants answered questionnaire) and recorded in SPSS v.23.

Descriptive analysis such as frequencies, means and modes of the responses to the categories gender, age, years of experience, and the rest of the closed questions that provided a Likert scale of responses, were run to find patterns among the answers and compare them among each other. In the case of the information recorded by each participant in the online questionnaire, the Isurvey platform of the University of Southampton provided automatically produced tables of frequencies with the data recorded. Therefore, this information was not put into SPSS v.23 software.

Answers to open-ended questions were examined separately to find patterns among answers of different participants. Answers to each question were classified according to the patterns found and then post-coded (provided a code after answers were given) according to the information given in the questionnaire. Information relevant to the research was delimited and a unique meaning was given to each code (Cohen *et. al*, 2011).

4.8.3 Ethical Considerations

As mentioned in the data collection procedures of this document, participants were explained the nature of their participation orally and in print through the Participant Information Sheet (Appendix M). This was provided during the training session one. This sheet included the objective of the study, the reasons why the participants were chosen, the risks and benefits of their participation, and the freedom they had of withdrawing from the

study at any time they considered it necessary. Finally, they were explained how their participation could be considered in only two phases of the study (phase one, two and three) or in the complete study (Phases four and five). Whether they decided to take part partially or fully in the study, the participant signed the informed consent included in Appendix N of this thesis.

In an interest to protect the identity of the EFL teachers, language managers, students and that of the participating institutions, coded IDs were provided to each participant and their institutions. Their personal information was not revealed and only data that was crucial for the study was analysed. To keep data confidential, the same assigned coded ID numbers were used throughout the multiple data collection stages with the purpose of tracking down any change in participants´ assessment. Data was stored in a password protected computer and in a password protected personal cloud storage software available online (Google Drive) with the purpose of avoiding data loss. Only the researcher had access to participant data.

With the purpose of diminishing the Hawthorne effect, the social desirability bias (Dörnyei, 2007, p.53) and avoid having participants provide answers to questions elicited during the interviews that do not represent their true experiences or `exhibit performance that is believed is expected from them´ (Ibid, p.54), a data triangulation method was implemented by involving multiple data collection instruments that could allow to come to the same conclusions. However, certain researcher bias may be involved in the results portrayed in this project and should be considered since it was the researcher who provided the assessment training sessions and conducted the pre-and post-training interviews. For

this reason, triangulating information with the online questionnaire, during which participants individually provided their perceptions, were also considered as triangulation procedures.

Finally, data obtained from the multiple data collection instruments was analysed by the researcher and then peer reviewed by an external experienced researcher in the field of applied linguistics (Dörnyei, 2007, p.61) with the purpose of comparing correspondence of results obtained.

With the use of all the instruments and the procedures previously detailed to analyse and collect data, specific results were obtained that could provide an answer to the five research questions that lead this study. The following Chapter exemplifies these results in depth.

Chapter 5: Results

This chapter outlines the results obtained from the analysis of the data depicted in the multiple data collection instruments used in this study. Semi-structured interviews to eleven teacher participants, four language managers and four groups of students were analysed following qualitative principles. Closed-ended answers provided to the background questionnaire, the online questionnaire, and teachers´ analytic and holistic scores to five written samples were considered for descriptive and inferential statistical analysis. The results are organized according to the research question they intend to answer.

5.1 Impact of Assessment Training on Reported Classroom Assessment

To answer RQ1 (*To what extent does writing assessment training impact EFL teachers´ reported classroom assessment of students' writing skills?)* data was obtained from the pre-and post-training interview transcripts of the participating teachers of phases four and five. Data found lead to the interpretation of the results outlined in this Chapter and the proposal of the Writing Assessment Training Impact Categorization (Figure 7, further described in Chapter 6 of this thesis.) which seeks to provide a classification of the effects that training triggered in participating teachers. As can be depicted in Table 6, data suggested that assessment training had an impact on writing in the EFL classroom, on teachers´ regular classroom assessment procedures of EFL writing, and on their self-awareness as an EFL teacher or as a classroom assessor.

`Writing in the EFL classroom´ was the first emerging main theme. Two subthemes were found among this main theme: A) writing activities and B) feedback techniques. Participating teachers reported to have implemented specific activities post to training that were categorized within these two subcategories. In regard to subtheme A (writing activities) teachers reported the 1) implementation of varied writing activities, 2) an increase of writing activities, and 3) implementation of activities to suit students´ needs. In regard to the second subtheme (feedback techniques) teacher participants pointed out that they had 1) implemented distinct feedback techniques, 2) increased their feedback to students and 3) improved the feedback provided to students.

The second aspect on which teachers reported training caused impact was `Classroom Assessment´ (Main Theme 2). It was mainly found to be characterized by two subthemes, A) Assessment Procedures and B) Scoring tools. Teachers explained they experienced distinct changes in their assessment. For instance, those that experienced impact in their assessment procedures (subtheme A) experienced 1) implementation of a new assessment technique such as student self-assessment; 2) an innovation of an assessment techniques (replaced a writing activity for a new one); and 3) a reorientation of their assessment purpose. Participants that pointed out they experienced impact in their use of Scoring Tools (subtheme B) found they 1) implemented a new scoring tool, 2) innovated their current scoring tool, 3) increased their use of the current tool and 4) combined the use of two tools depending on their purposes.

Writing assessment training was also found to trigger `Teacher Self-Awareness´ of their current assessment practices and other activities in their classroom (Main Theme 3). For

instance, teachers reported they reflected about the Nature of Writing (subtheme A), the Teaching of Writing (subtheme B) and, their Assessment Procedures (subtheme C). Meta cognitive analysis was found to also trigger teachers´ Writer Stance (subtheme D) and their Student Stance (subtheme E). Table 6 graphically depicts these main themes and subthemes as well as the subcategories described.

Two teachers denied to have experienced direct impact of training in their language classroom. Specifically, teacher participant (TP) 16 and TP326. One of these participants, TP16, explained that although he did not implement change, he noticed he had reflected on his current assessment practice which allowed him to prepare future assessment procedures. TP326 reported to have not implemented change due to the small amount of class time and added,

> …but sometimes, because of time constraints, we cannot rely on the rubric, we know what it is, or because of experience we know what is a good piece of work or which is a regular piece of work, which is bad, so we only do it according to our memory or practical experience.

The same reflection was reported by TP23, TP32, TP37, and TP325. These TPs indicated that their assessment in the classroom did not receive direct impact. However, they specified that the teaching of writing had increased, therefore being considered as part of Main Theme 1.

Table 6 Impact of Assessment Training on Participants of Phases 4 and 5

Main Theme	Subtheme	Teacher Participant
Impact on Classroom Teaching	Writing Activities	TP23, TP37, TP315, TP325
	Feedback Techniques	TP23, TP32, TP37, TP325
Impact on Classroom Assessment	Assessment Procedures	TP22, TP62, TP313
	Scoring Tools	TP22, TP62, TP73, TP313, TP315
Impact on Teacher Self-Awareness	The Nature of Writing	TP315, TP325
	The Teaching Writing	TP23, TP37, TP313, TP315, TP325
	Assessment Procedures	TP16, TP22, TP23, TP32, TP62, TP325
	Writer Stance	TP23, TP32, TP37, TP315
	Student Stance	TP315, 325
No impact was experienced		TP16, TP326

The following section describes participants´ reported impact of training through the interpretation of the information obtained from the pre and post training interviews. It begins with a description of data obtained from the pre-training interview and is followed by the responses that teachers provided to the post training interview. These responses were then compared and contrasted to compile the main levels, sub levels and sub categories of the Writing Assessment Training Impact Categorization (WATIC, Figure 7) which is further described and explained in the Discussion (Chapter 6) Chapter of this thesis.

5.1.1 Teachers´ Pre-training reported Teaching of Writing and Assessment Procedures

As portrayed in Table 7, data obtained from the pre-training interview revealed that TPs' reported teaching and assessment practices in their classroom corresponded to three main themes: 1) TPs that gave a high level of importance to the teaching and assessment of writing in their practice, 2) TPs that gave a minimum level of importance to the teaching and assessment of writing for distinct reasons and 3) those that did not teach or assess writing in their practice. Category 2 was the most common among the reported practices (TP 22, TP23, TP32, TP37, TP62, TP73, TP313 and TP315) while Category 3 was the least common with one TP reporting this practice, TP325.

For instance, participants who believed that writing was very important and gave a high level of importance to its assessment, in terms of the percentage of scoring criteria, were classified in Category 1. Three participants resulted to be classified in this category because they either engaged their students in many writing activities and/or gave a high percentage to writing in their regular classroom assessment. However, these TPs´ pointed out that the most common issues they faced when assessing writing in their classroom were `program issues ´ and ` time issues´ (as signalled in Table 7). In other words, they struggled to comply with the requirements of the EFL program they taught and its focus on other skills that were not writing or they simply did not have enough time in the classroom to dedicate to writing and its assessment.

The second Category (2) `Minimum Level of Importance to the Teaching and Assessment of Writing´ is characterized by the little importance the participant gave to the skill. They

asked their students to develop writing tasks without considering it part of their regular assessment or giving it a minimum percentage (5-10%) of students´ final grade. Seven teachers were found to correspond to this category and stated they also faced issues like Category 1 teachers but added a new type of problem of their own ` Particular Issues´. In other words, these teachers believed their lack of organizational skills were a constraint to properly dedicate time to the assessment of writing. Therefore, stating that their own particular teaching and assessment style needed to be improved.

Finally, TP313 and TP325 signalled that they did not teach or assess writing (Category 3) because the load of work dedicated to the development of other skills in their EFL program did not allow them to teach and assess writing in their classrooms. Therefore, pointing out they gave a high level of importance to skills such as speaking, reading and subskills such as vocabulary and grammar.

Table 7 Teacher Participants' Pre-Training Reported Teaching and Assessment Issues

Category	Teacher Participant	Issues Faced
A	TP62	
	TP326	Program Issues
B	TP22	
	TP23	Program Issues
	TP32	Time Issues
	TP37	
	TP315	
B	TP32	Teachers´ Particular Issues
	TP62	
C	TP313	Program Issues
	TP325	

A: High level of importance to teaching and assessment of writing. **B:** Minimum level of importance to the teaching and assessment of writing. **C:** No assessment or teaching of writing

Although these participants stated to follow specific assessment procedures prior to experiencing training, some of the TPs explained that their perception and regular assessment practice changed on some aspects post to training. These aspects are further described in the following sections.

5.1.2 Teachers´ Post-Training Reported Teaching of Writing and Assessment Procedures

As previously described, training was found to impact three main areas 1) the teaching of writing in the EFL classroom, 2) writing assessment in the EFL classroom and 2) teachers´ self-awareness of their teaching and assessment processes. The following sections provide a description of the findings that lead to the categorisation of the impact of writing assessment training.

5.1.2.1 Teaching Writing in the EFL Classroom

The first main theme found among post training impact was innovation to the teaching of writing in the EFL classroom. Among this theme, two subthemes emerged: a) Writing Activities and b) Feedback Techniques.

a) Writing Activities. Within the first subtheme, TP23, an experienced EFL teacher, reported she continued giving importance to the assessment of other skills mainly because of time issues and EFL program demands. Post to training she reflected on the importance that writing has for language students and their need of it in their future professional lives. She explained that she now includes more writing activities in her lessons and provides

more feedback to students' texts with the aid of a correction code as pointed out in the following excerpt,

> I implemented more writing exercises and I am using a correction code to provide students the feedback. I used to use a code but I only used two or three symbols and did not really give extended feedback…. I am trying to focus more and use it more…

From institution A, TP325 was who reported the most amount of impact after experiencing assessment training. During Interview 2, she reported to have more interest in writing and its treatment in her classroom. She explained that she now saw its importance in students´ language development resulting in her attempt to have her students write at least to a minimum level in the classroom or for homework (considering they did not write prior to training). She explained, ` …there is more interest from me in the sense not to leave it out…I started to put a little more emphasis on writing by writing at least a little or for homework depending on my students´ needs´.

It may be argued from this evidence that three categories emerged from this subtheme that reflect teachers´ changes in their classroom post to training, 1) Implementation of writing techniques, an 2) Increase of writing activities such as the case of TP23 and 3) Focus of activities on students´ needs carried out by TP325 as portrayed in the previous interview excerpt. These categories are portrayed in the Writing Assessment Training Impact Categorization (WATIC, Figure 7 p.237). Some teachers also managed to impact their feedback to writing which is described below.

b) Feedback Techniques. A second subtheme that emerged was the provision of feedback through different feedback techniques. Very similarly to TP23, participant 32, a male experienced teacher, pointed out that his assessment had not changed. It continued to be limited and without the use of a scoring tool. It was detailed that he read the text and gave it comments without considering it for students' monthly evaluation. However, he described his feedback had changed after attending the assessment training. Now, he was more careful and precise in the comments he provided his students. He became aware of the importance of feedback in students´ development of skills and in their assessment. He pointed out,

> I think my situation continues in the same tone. I still need more time to assess writing the way I would like to, I blame my disorganization with my time. I only read the text and provide comments…I believe that in the new methods to assess students´ feedback it is very important because if I tell the student "you failed" but I don´t say in what he failed or how he can improve then assessment would be useless we would only be giving a score

TP325, in addition to increasing the number of writing activities done in the classroom post to training, she modified her feedback focus by paying attention to the genre and the structure of the text students were developing. She paid more attention to the type of feedback she gave to her students specifically in the genre being taught as is explained in the excerpt below,

> I started to give more feedback in the sense of how they were basic level obviously and had much errors in their writing and how I needed to give more suggestions in their writing and use of grammatical structures…focused a little more on the formality because they tend to write casually and colloquially like that translated from Spanish to English… and I was giving more advice and support in that part…

Considering these previous comments and others received by those interviewed, it can be stated that three categories emerged from teachers´ activities post to training: 1)

Implementation of varied feedback techniques, 2) Increase of feedback provision and 3) Improvement of feedback provision. These are portrayed in the WATIC which is included in Chapter 6. More significant changes were reported post to training concerning teachers´ assessment procedures. The following section describes these in detail.

5.1.2.2 Classroom Assessment of EFL Writing

The second major theme that emerged from the transcript analysis was innovation in the assessment of writing in the EFL classroom. Specifically, two subthemes were identified in this main theme. Impact was found in a) the procedures followed to assess writing and b) their use of scoring tools to assess writing. The following section describes these subthemes and the participants that reported them.

a) Assessment Procedures. In subtheme 1, impact on classroom assessment procedures, TP22, an experienced 24-year-old male teacher considered that he gave a minimum amount of importance to the assessment of writing because the EFL program he worked with did not consider it an essential skill to develop in students. He only assigned 10% of the total monthly grade to written tasks. This participant reported he modified his leniency when assessing, his use of holistic rubrics to manage his time and his reflection of what he used to do when he was a more inexperienced teacher in comparison to what he does now as a more mature teacher. He specified that he used to expect more from his students than they could actually produce by stating,

> At some point, I had been too strict with my students and sometimes I would look at them and then interpret them without looking at their writing… especially when I am expecting something from them I was perhaps demanding the proficiency of level V in level IV. The sessions helped me notice that.

He explained that the training sessions helped him understand that at times it is necessary to demand from students more so they can perform better, but it needs to correspond to their level of proficiency.

TP62 was a female teacher who reported implementing small changes to her assessment procedures when focusing on her students´ texts. She explained that prior to the training sessions she would read her students´ work, directly provide the corrections to the text and score it according to her personal judgment. Scores were given to students without any type of feedback on how to improve their writing skills or any prior explanation of what would be assessed on their work. She pointed out that post to training she had been able to implement change to her regular procedure by explaining to students prior to the assessment the scoring tool used. She explained that,

> … at first it was merely my judgment: I read it, corrected it. I would not let them do so but now …I looked for a tool that fits their level and gave it to them before I applied the writing task…I actually read their work again and never gave them feedback. I corrected them, crossed it out and did not give them the opportunity to reflect on what they thought they were doing.

TP313, explained during the second interview that the two training sessions had helped him implement change in his assessment techniques in the classroom. He first explained that prior to the training he did not consider encouraging students' self-assessment of writing. Post to training, he had managed to implement it with the help of a self-correction code. He considered that this implementation had resulted in an increase in students' awareness of the importance of the assessment of writing. He pointed out,

> …with that group, I was able to notice that before students did not have a clue of how to evaluate their own work they relied completely on the teacher. After the training, I was able to implement techniques of how to evaluate each other and they were able to understand the use of a different evaluation…they began paying more attention to things they did not know and to pay more attention to their self and teacher evaluation…

After his implementation, he found students had learned to pay more attention to their work while also being interested in figuring out the meaning of the symbols of the correction code.

Considering these TPs´ reported training impact, it can be argued that specific categories emerged from the subtheme `Assessment Procedures´. For instance, TP22 suggested he had a) Reoriented his assessment purpose, TP62 reported she had b) Increased her provision of assessment feedback while TP313 explained an c) Increase of the use of assessment techniques. These can be further visualized in Figure 7 (WATIC, Chapter 6 p.237).

b) Scoring Tool Use. In regard to subtheme 2, impact on scoring tool use, TP22 reported to have changed his scoring tool post to training sessions. He pointed out that he had always used an analytical scoring tool with all his students regardless of their interests, abilities or needs. He now considers their proficiency (the lower the proficiency the more general the scoring tool) and his purpose when assessing students´ written work. Therefore, this participant shifted to a holistic approach to assessment to provide a more meaningful score to their work and for managing his time more wisely. He explained,

> I pay a little more of attention to the rubrics… I use rubrics all the time depends on the level and it depends on the type of writing I am checking. I am more holistic now it depends on the task …I am trying to be more holistic because it is a lower level and because I have more students and I need to administer my time… I went back to holistic to have it either on my screen or to have it next to me on paper and give a valid and reliable result for every task my students elaborate…

In the same public institution as TP22, TP62 explained that prior to the training she used an analytic scoring tool that she then found did not suit the capabilities of her students´. She described how she looked for a rubric on the Internet and adapted it to her assessment process to suit her students´ proficiency and her own assessment purposes in the classroom. She also sought to implement a correction code as a tool to encourage students´ reflection, self-assessment of their texts and described,

> …I looked for a rubric to fit their level and gave it to them before I applied the writing task I also gave them a code and now I don´t correct their work I use the codes so they can self-evaluate their work. They improve their text and then I give them the score…now I am asking them the original with their corrections and the final version of their text. It has had an impact because they ask me things like "What does this code mean?", "How can I improve it?" they are showing more interest. They do once more and they return their final draft to me with the initial draft with my feedback. The final draft then receives the score.

During Interview 2 TP73, an experienced female teacher, explained that although she did not have the opportunity to change her assessment methods fully throughout the semester, she did experiment a single task with her students. She combined feedback techniques with the use of an adapted scoring tool to focus her assessment on aspects students needed to improve. She pointed out that after the training sessions, she focused her assessment on

specific aspects that students needed to develop and used an analytic scoring tool she adapted from other sources. She mentioned,

> After the sessions, we had one more writing to work on before the period finalized. I assessed it differently. I used to use a correction code to give feedback on accuracy and written comments for content purposes. But this last text, I experimented focusing on more specific aspects yeah like format, use of linking words, transitions. I adapted a very simple rubric, very simple that I got from different sources on the internet. It included the most common categories and I added one so I could focus on transitions. I showed students the rubric and I explained what I expected from them. They seemed like they understood.

She specified that she considered the training sessions had allowed her to adapt a different tool to her classroom needs. However, she considered that this cannot be done in every class because writing assessment, assessment tools and assessment purposes depend on the EFL program´s goals and students´ proficiency.

In relation to analytic and holistic scoring tools, TP313 specified that after taking both training sessions he had been able to combine the use of both types of rubrics, analytic and holistic, not just to assess students' written work but also as a tool to provide feedback to them. He continues using the analytic rubric as he did previous to the sessions however he integrated it into his regular teaching and assessment activities. He used it to guide students so they know what is expected from them and as a resource to identify areas that students need to improve. He indicated that he found ways of managing his time by integrating a holistic rubric to assess students´ performance on the monthly test and the analytic rubric as a feedback tool. He explains this in the excerpt below,

> …Yes now I use a holistic rubric because it results easier to use because of the time and the amount of students. I used a rubric only to evaluate tests but I was able to adapt them to all my activities, I still use rubrics and I implemented them to the development of the students more than anything to their development and the great majority seemed interested and they didn't get confused…I think it is better and you maintain better control over students' evaluation because without something to guide you of how to evaluate of how to classify mistakes it's kind of hard to define how to help them.

This change in conceptualization of rubric and its use allowed this participant to improve his assessment procedures in his classroom, which resulted in students´ easier understanding of the task and a smoother development of writing skills. He reports this perception by pointing out,

> I felt that with the analytic rubric they had a better idea and they said to me "yes teacher it is easier I can see step by step". They needed something clear…that could explain to them what to do.

Regarding the use of analytic and holistic rubrics to assess writing, she commented that before taking the training she did not know they were available to facilitate teachers' assessment practice thus did not use any type of tool. Post to the training, she understood how a rubric could be used to standardize classroom assessment of writing therefore allowing it to be more objective and valid for students. She added that these tools provided to her by the researcher were initially used in her class without any type of adaptation and then she gradually found a way to adapt them to her students' needs. Thus, implementing rubrics to assess students´ classroom writing and provide feedback to her students. She commented,

> ...Yes actually a lot...I use the rubrics that she (the researcher) gave us, I was using them exactly the same and then for this course I adapted them, each activity is

> different, everything I ask them is different, the tests ask for different things, then the rubrics are adapted. I use them very generally and then I only add things…

Finally, TP313 added that assessment training had also allowed her to feel more confident in her assessment procedures and less "dirty" when assigning a score. It was explained that the rubrics allowed her to have an objective explanation to a specific grade given to students' texts therefore setting aside her personal views. She explained this by stating `I don't feel "dirty" anymore every time…I have the base established of why you obtained a 9…because of this and this (signalling the rubric) here it is, here is what I did´.

Considering the comments provided by TPs in relation to this second subtheme (Scoring Tool Use) corresponding to the second emerging main theme (Writing Assessment in the EFL Classroom), it can be argued that four subcategories emerged: 1) Innovation of current scoring tool (exemplified with TP22), 2) Implementation of scoring tool (portrayed by TP62), 3) Adaptation of current scoring tool (pointed out by TP73) and a 4) Combination of scoring tools (reported by TP313). These emerging categories can be visualized in the WATIC (Figure 7, Chapter 6 p.237).

5.1.2.3 EFL Teachers' Self-Awareness

The third major theme that emerged was impact on teachers´ metacognitive skills. For instance, teachers reflected (emerging subthemes) on 1) the nature of writing, its 2) teaching of writing in the EFL classroom and 3) the procedures followed to assess it. It was also found that training triggered TPs´ self-reflection of 4) their stance as a writer and as 5)

a student. The following section describes the five emerging subthemes and the participants that reported them.

a) The Nature of Writing. This first subtheme focuses on those teachers that reflected on the importance of writing as a language skill post to training.

TP325 experienced a change in her view of writing and its importance in students` language development. She commented that she tried to help her students change their view of writing as a difficult and unachievable skill by changing her own view. She pointed out,

> …my job is to make them see reality and change that perspective and it is difficult to change them (the students) …and yet I have now started having them see an easy way of writing and …if I change my mentality that is something I need to do in the classroom I need to give time to it (writing) and find a way to do it and give it a little time for feedback

TP37, an experienced female teacher added that at the personal level the training sessions had allowed her to recall the importance of writing for students and for their future professional life. She explained how she had always known of this importance but chose to work on other things. She explained how she felt by stating `… I feel it (assessment training) helped me remember things: to give writing the importance it should have in my classroom, despite of the time issues…´. Finally, TP313 explained how she had reflected on how writing is an activity that is best learnt as a social activity and stated `…I was able to see how writing works better when shared with someone…´.

Considering the views of these participants, within subtheme one (Nature of Writing), and as shown on the WATIC (Figure 7), two categories emerged. For instance, a) Importance of writing for a language student was reported by TP325 and TP37 while category b) social role of learning to write was exemplified by TP313. This leads to the second emerging subtheme, which is described in the following section.

b) The Teaching of Writing. In regard to teachers´ reflection of the treatment of writing in the classroom (subtheme 2), TP23 pointed out that his assessment methods did not change post to training sessions. However, he explained that he did have an opportunity to analyse how he was teaching his students writing and the little importance he was acknowledging to the skill. He clarified that he was now aware that assessment could be aided with the use of a tool to standardize its assessment. But he needed to give more importance first to its teaching then move on to its assessment. He found his lack of organizational skills another factor that affected his lack of change in the classroom. The excerpt included below depicts these perceptions,

> … something I would rescue is that I thought a lot about what I was doing in the classroom and as an EFL teacher…I think I also had the opportunity to become aware of how disorganized I am with my time. I want to sit down and organize my time and my activities so they are not just another activity…

TP37, explained that she had not changed her assessment process in the classroom because she felt unprepared. She explained was still finding ways of increasing writing activities in the classroom. This TP pointed out that after the training sessions, she understood that her current techniques to teach writing needed to be improved. She added that she increased

her writing activities in the classroom even though it required large amounts of time. She explained that,

> …more than assessing I continue asking them to write…This training has helped me to avoid being careless about my teaching of writing despite of the lack of time we have to finish the program. We can never finish it, we can´t even finish the minimum. Writing takes a lot of time and homework for the teacher, and despite of that I have tried to make the effort to give it more time in the classroom so that students can practice…. I implemented three activities from which one was done for homework and the rest were done here in class.

This second subtheme focused on the Teaching of Writing, which also reflected emerging categories in teachers´ reflection. Categories identified were a) Improvement of teaching skills, b) Future inclusion of writing, c) Future inclusion of feedback and d) Future inclusion of process writing. These subcategories were reported by TP23, TP37, TP313, TP315 and TP325 and may be found on the WATIC (Figure 7, Chapter 6 p.237).

c) Writing Assessment Procedures. In relation to participants´ reflection of assessment and its process in the classroom (subtheme 3 identified within the main theme `Teacher Self-awareness´), TP23 reported to continue assessing writing in her classroom to a minimum level without a specific procedure being followed. She explained she was still analysing a possible way of implementing scoring tools in her classroom assessment. She described how the rubrics provided by the researcher were overwhelming and difficult to use. It was explained that she only focused on her students doing what was required and communicating meaning to the reader. She detailed,

> … to use one (rubric) it needs to be which best suits you for example the ones she gave to us I found them very heavy. It was a lot of information then to be checking the activity and reading the column and deciding, it was too much…Depending on the purpose of the activity is what I evaluated for example if he had to communicate

> my daily routine if they managed to do so they succeeded, fine. If the goal was that they express their ideas using the present continuous and they did fine.

She finalized this comment by expressing that she understood rubrics were useful but she did not use them because she did not know how to do so. Therefore, preferring a checklist to assess and in which it was recorded if the student fulfilled or not an activity.

In relation to scoring tools, TP32 explained he had noticed the need to implement tools to assess his students' writing and explained he would like to implement an instrument in his lessons preferring an assessment checklist. He considered it would be easier to use in the classroom. TP32 specified he did not change his classroom assessment of writing post to assessment training. However, as with TP23 the sessions allowed him to reflect on his present activities in the classroom.

A second participant that reflected on his use of scoring tools was TP22. He reported he considered that one of the major gains he obtained from the training sessions was recalling his experience when learning to assess languages in his initial teaching years. He considered that teachers need to be constantly updating their teaching and language skills. As a novice teacher, much input and suggestions are received but when an experienced teacher is exposed to training it provides them opportunity for recalling and refreshing information that can improve lessons and content provided to students. In this sense, TP22 explained his view of the impact of assessment training on his use of scoring tools and its improvement by stating,

> …but eventually you go forward and become a teacher and you tend to forget that students at some point of the class are bound to feel confused… the same thing happened to me with rubrics I used them at the beginning of my teaching very frequently for every activity and then I kind of memorized the process and it became a habit and I feel I got stuck doing things the same way all over again and I wasn't able to adapt to my classes because I had a veil over my eyes which was implemented by routine. The (training) sessions helped me remember that feeling like ohhhhh my students are different people… I went back to using rubrics.

In relation to the regular assessment procedures followed, TP62 explained that training had helped her reflect on what she was doing in her classroom, how she was doing it, and therefore plan how she could improve her assessment to make it an easier task for her and more reflective for her students. Therefore, updating her assessment skills. It was pointed out that she had been able to feel more confident and more objective in her explanations to students' doubts about their scores as depicted in the excerpt below,

> I even had someone ask me "Why did I get this score?" and I answered "Check the scoring scale and comments I gave you. Check what you got and analyse it and if you still have questions come and tell me". I plan to continue… like this because it is easier for me even if I have a lot of students.

Also with regard to assessment procedures followed, TP73 pointed out that, she had had the opportunity to reflect on how she was doing things in her classroom. She explained that many of the times teachers get caught up in their routines or their lack of time and training can help teachers recall assessment procedures and tools that are available that can improve and facilitate their job. The excerpt included below depicts these ideas.

> It is really good to get these kind of experiences (teacher training). It is not that we do not know…well sometimes we do not know…but these sessions allow us to remember. Because sometimes we get into a routine like "I do so because it is how I manage my time". We are used to using one all the time and sometimes it is not good because assessment tools should change depending on text type and we need to recall which tools we have available. So, in my case, yes training helped me remember things I had forgotten...

TP16 explained that he had not had the sufficient time to implement change in his classroom assessment after the training sessions. He stated he continued assessing students' written work with their portfolio work and with a monthly exam that included a writing component. However, he reported that the training had given him the opportunity to reflect on how their actual assessment methods could be improved. He reflected on how assessment tools were being combined and how he considered the use of the portfolio could allow students´ development and reflection, as he states in the following excerpt,

> We did not have a chance to implement… I´m thinking a little bit more on changing the way we evaluate students, I consider of course the portfolio is an important part because you are evaluating students continuously, and umm in the exam for example, you have four exams you fail one exam, you cannot, you can do nothing about the grade, you cannot say ok if you do it next time better I will give you a better grade for the first part of the first exam, it is not possible the grade is there, and it is not possible. With portfolio work you are having products, and you are making them better, it is a better way to evaluate because you are learning from your mistakes for example.

This participant explained that training sessions had allowed him to think about the use of the analytical tools to score writing. The institute at which he worked has been providing the same analytical rubric to assess writing from Introductory Levels to Level Ten for more than eight years. Therefore, taking the training session had enabled him to reflect on how a change was necessary. He stated,

> We (teacher staff) spoke mainly about the writing rubrics if we should change them or not here at CELLAP many believed it is too general, the rubric is used for every level so there is no rubric for each level. The session… guided us to find new ideas… In the fall semester, we are going to have three academic sessions and one of the sessions we are going to talk … about changing the rubrics, making rubrics for each level, collaborate together to make them.

Participant 315 was a young novice teacher who was interviewed and explained how she initially had issues combining formative and summative assessment in the way Mexican teachers are required from their institutions. In other words, quantifying what students know about the target language. She explained that the assessment training sessions had allowed her to feel more confident in her scores and secure when students, parents or the administration of the institute required further explanations. The following excerpt explains this,

> … yes it totally changed the way I saw how grades are given…I have always had an issue in giving a number to how you are learning a language like from 1 to 10 how much English do you know I find it illogical and I have always had an issue with that... at the end of the day I have to assign a number and I now understand that it is part of teaching a class in any institution…

Finally, she expressed she now felt more confident and secure about the score she was giving the paper. She manifested she had found a way of combining the institutional requirements with her own assessment beliefs.

Participant 325 signalled that she did not implement change in her assessment procedures with her actual students. However, the training had allowed her to begin planning her future assessment of writing, in relation to the purpose of assessment as well as planning

the future assessment workshops she would like to attend. Specifically, what techniques and tools she was going to use to fulfil her students' needs. She specified that she was currently analysing what type of scoring tool to use to avoid having her students stress out and instead allow them to improve their writing through assessment. She stated,

> I had no opportunity… to implement it (assessment of writing). However, as I will continue working with the same level I am still working on an adequate tool for the group because it is a very basic level that they almost do not know how to write. They need to know more vocabulary, …I want it to be less stressful for them I still have not decided, …I will not be very rigid with them I don´t want it to be frustrating for them, without feeling worried, I want them to see slowly that they do know that they can do it and that they can start little by little…

On the other hand, TP37 pointed out that she continues without using a scoring tool to grade the final writing project because she lacked the knowledge to use it and she considered her students did not have the skills to understand what it meant. She justified her decision of not implementing a tool at this point until the students and her were prepared. The following excerpt depicts this information.

> I haven´t been able to implement any type of scoring tool, I don´t feel prepared and they are not ready. I feel I would frustrate them, I would feel frustrated…but maybe next semester with more time they will be ready…To use rubrics with their work I need to prepare them and little by little let them go

These excerpts may suggest that writing assessment training (WAT) allowed teachers to reflect on the procedures they followed to assess their students´ work. To explain this and to point out the categories that emerged, TPs reported to have 1) Updated their assessment techniques (reported by TP16), 2) Updated their assessment procedure (exemplified by

TP22 and TP73), 3) Began planning of future assessment (portrayed by TP62 and TP32), 4) Corresponded teaching and assessment purpose (pointed out by TP325) and 5) Considered students´ needs (explained by TP37). These and the previous categories identified can be found in the WATIC (Figure 7, Chapter 6) on page 237 of this thesis. TPs also reported to have analysed their abilities as a writer of English. This is further explained below.

d) Writer Stance. A fourth subtheme that emerged as part of Main Theme Three (Teacher Self-Awareness) was participants´ conceptualization of themselves as writers. TP23 reflected on the need for her to write to therefore transmit to students the skills needed to develop a text. She became more conscious of her weaknesses and her needs as a novice writer. She explains so in the following excerpt.

> …I´ve become more conscious that it is a skill we need to teach and evaluate. But, as an English teacher, writing is a skill I am deficient, I'm not good at writing so to be able to teach you need to know how to do it.

Another participant that expressed the training had allowed her to reflect on herself as a writer, was TP37. She explained that post to the sessions she had been able to analyse herself and conclude that she had weaknesses that needed to be improved, and if improved there would be a possibility of providing more quality feedback and assessment of writing in the classroom. In this sense she stated,

> …it helped me understand that I also need to work with my writing, we need to feel with confidence in our writing…if I learn to improve my writing it might be easier to improve my students´ writing.

TP23 and TP37 both stated that their reflection on their writing weaknesses led them to visualize their needs to be further trained began planning to seek for other courses or workshops to attend that could allow them to improve their writing skills and their professional development.

It may be concluded that TPs analysed and reflected on their performance as writers therefore identifying the following categories in teachers´ reflection: a) Weaknesses as a writer (TP 23 and TP37), b) Improvement of teacher writing to improve student writing (TP 37) and c) Strengths as a writer (TP32 and TP315). The WATIC, included in this thesis as Figure 7 in Chapter 6, maps out these and other emerging themes from this project.

e) Student Stance. Teacher participants reported to have reflected on themselves as students who are constantly being evaluated (subtheme 5) in their programs of study and in their working environment. For instance, TP315 explained that writing assessment training had helped her in different ways. Firstly, to understand what to consider when assessing her students and when being assessed by her BA professors. Secondly it had allowed her to better understand the use of scoring tools, and/or to adapt them to her needs and students' needs.

During the interview, TP315 explained she had changed her perspective as a student and as a teacher about assessment and all the factors that have an active role in it. Before taking the assessment training she did not consider what her students were being taught in the classroom but instead only focused on the quality of a product. She pointed out,

> …but also, I advanced personally, as a teacher and student I 've realized that you cannot isolate writing, then I found a way to balance, you need input to produce output, I cannot evaluate only what you are giving me, so I mean that I'm giving you input that I have to take into account…so I think my professors did not only evaluate what I wrote… but what I understood of what they taught…

With this excerpt, it may be inferred that as a student she felt more at ease with her professors´ assessment and as an in-service teacher she grew as a professional by gaining a deeper understanding of writing assessment. It can be considered that as a BA student she was able to further understand how her professors connect classroom activities with assessment tools. Corresponding to this view, TP325 explained that as an MA student she had a difficult time understanding her professors´ assessment procedures by explaining

> …it seemed my professors´ were against me but after the training I remembered some of their explanations as to why I had gotten a specific grade…now I get what they tried to explain…

It may be argued that this TP, additionally to the reflection gained about writing assessment, a deeper understanding of her student performance and her professors´ assessment was understood.

From this data in can be inferred that participants TP315 and TP325 were impacted in their stance as a student in their 1) understanding of assessment knowledge and became aware

of 2) their performance as a student while being a BA or Master´s student. These subcategories are portrayed on the WATIC (Figure 7, Chapter 6).

5.1.3 Section Conclusion

This section focused on giving response to the first research question that lead this study (*To what extent does writing assessment training impact EFL teachers´ reported classroom assessment of students' writing skills?).* Results suggested that the impact on teachers´ actual classroom assessment was observed on few occasions. Those who managed to innovate their assessment process and their use of scoring tools reported to have done so on a single occasion during the study. Nine out of the eleven participants of phases 4 and 5 reported to have experienced positive impact in their assessment, and/or their teaching of writing. Assessment training was found to most frequently impact other aspects of teachers´ growth as teaching professionals and classroom assessors, such as their meta-cognitive skills in which they reflected on themselves as writers, as English teachers and as assessors who take an active part in an institution and its assessment policies.

During data collection, teachers also provided their perceptions about the training sessions and about writing assessment itself. The following section focuses on these perceptions and how they changed throughout the study.

5.2 Impact of Assessment Training on Teachers´ Perceptions of Writing Assessment

To answer RQ2 (*What is the impact of assessment training on teachers´ perceptions of writing and on their perceptions of classroom writing assessment?*) semi-structured

interviews (one conducted prior and one post to training) and a post training online questionnaire, were used as instruments to collect this information. The data obtained from these instruments suggested participants´ perceptions changed but not only those regarding to the assessment of writing, but also of writing itself and its importance. The following section focuses on the perceptions of the forty-eight participants of Phase One, Two and Three of this study and the eleven participants who actively participated in Phases Four and Five.

As shown on Table 8, the open and closed questions included in the post training online questionnaire were oriented towards three main topics: a) participants' perceptions of assessment training, b) impact on their use of scoring tools and c) participants´ performance during the study. The closed questions provided respondents with a Likert scale (totally agree, agree, neither agree nor disagree, disagree and totally disagree) to rate their perceptions while the open questions allowed participants to provide their perceptions freely.

Responses to the questionnaire suggest that a higher percentage of people felt more positive towards the assessment training sessions in comparison to their use of the provided rubrics. In other words, a lower number of teachers felt that the use of scoring tools such as analytic and holistic tools did not become easier, or more efficient to use. A little over half of the participants considered the rubrics shared by the researcher/trainer useful for their actual assessment. Therefore, suggesting that, globally, participants did not feel comfortable with the use of scoring tools.

Teacher Participants did consider that the content of the training was clear, understandable, useful and practical. This may seem rather contradictory since the main goal of assessment training is to facilitate assessment processes and teachers' use of assessment tools. However, getting familiarized with scoring tools may take more than two training sessions due to their complex nature and the subjectivity of writing assessment.

When questioned about the training content of their preference, teachers suggested that it should include more practical experience assessing texts instead of the trainers´ presentation of theoretical concepts. This may suggest that teachers are well aware of their needs and flaws in regard to assessment and seek for practical and useful sessions.

5.2.1 Teachers´ Perceptions of Training Sessions

As shown on Table 8, the majority of the TPs totally agreed that the content of the session had been clear and understandable. While 89.6% of the TPs chose `strongly agree´ when suggested the training had been clear and understandable, 8.3% chose `agree´ and 2.1% chose `strongly disagree´ (one participant). Participants were also elicited suggestions regarding training content and their perceptions about what writing assessment training sessions should approach. The aspect that was the most favoured was `Discussion of the distinct types of rubrics and their use´ (87.5% of votes) while the least favoured was `Theoretical background to the evaluation of writing´ (20.8% of votes).

Table 8 Teachers´ Perceptions of Assessment Training: Online Questionnaire

Statement	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
Perceptions of Training Sessions					
1. The information and practice shared during the training session was clear and understandable.	89.6	8.3	0	0	2.1
2. The information and practice shared during the training session is practical for my future evaluation of students´ writing.	85.4	10.4	2.1	0	2.1
3. The information and practice shared during the training session is useful for my future evaluation of students´ writing.	83.3	8.3	4.2	2.1	2.1
Impact on the Use of Scoring Tools					
4. After taking the training session, I consider that my use of rubrics has become more efficient.	66.7	22.9	6.3	2.1	2.1
5. After taking the training session, I consider that my use of rubrics has become easier.	52.1	37.5	6.3	2.1	2.1
6. After taking the training session, I have decided to use an evaluation tool such as a rubric to assess my students´ writing skills.	54.2	33.3	8.3	2.1	2.1
Participants´ Performance in Study					
7. The rubrics provided by the researcher/trainer will be useful for my future evaluations of writing.	70.8	20.8	4.2	2.1	2.1
8. After the training session, the scoring of the writing samples provided by the researcher was easier.	64.6	22.9	8.3	2.1	2.1
9. After the training session, the scoring of the writing samples provided by the researcher was more efficient.	68.8	20.8	6.3	2.1	2.1

Regarding the practicality and usefulness of training for TPs´ future assessment practice. Once again, the majority of the TPs strongly agreed (85.4%) with its practicality while five (10.4%) TPs agreed, one (2.1%) neither agreed nor disagreed and one (2.1%) totally disagreed. In relation to its usefulness, 83.3% of TPs totally agreed to the statement `The information and practice shared during the training session is useful for my future evaluation of students´ writing´, four agreed, two neither agreed nor disagreed, one disagreed and one totally disagreed.

5.2.2 Impact on the Use of Scoring Tools

When elicited their perception towards the changes in their use of rubrics, the majority of the participants perceived changes had arisen post to training session provided. More than half of the TPs (66%) totally agreed their use of rubrics had become more efficient after assessment training, 22.9% agreed with this statement, 6.3% neither agreed nor disagreed, 2.1% disagreed and another 2.1% totally disagreed. To the statement `After taking the training session, I consider that my use of rubrics has become easier´, 52.1% TPs strongly agreed, 37.5% agreed, 6.3% neither agreed nor disagreed, 2.1% disagreed and another 2.1% strongly disagreed.

In an effort to be more specific regarding the changes they experienced in their use of scoring tools TP04, TP05, TP12, TP22, TP26, and TP34 explained in open question eleven that their use of the tools became more objective, post to training while TP73, TP28, TP27, TP32, TP42, TP319, TP303, TP312, TP317, TP314, and TP311 considered their assessment improved after training. Other explanations encountered were that training made use of scoring tools easier, more agile and more useful. Two participants (TP14 and TP315) considered their use of tools did not change after receiving training stating the rubrics provided were very similar to those used in TP14´s workplace while TP315 considered training was `more useful to measure use of rubrics´.

Another change elicited was TPs future use of scoring tools. From a total of 48 participants, 54.2% TPs strongly agreed to the statement `after taking the training session, I have decided to use an evaluation tool such as a rubric to assess my students´ writing

skills´, 33.3% of the TPs agreed, 8.3% neither agreed or disagreed, 2.1% disagreed and another 2.1% strongly disagreed.

When eliciting participants´ preference of rubrics, 52.1% of the TPs mentioned that they preferred the use of a holistic and analytic scoring scale after taking assessment training while 41.7% preferred only using an analytic scoring scale. Only 6.3% of the TPs preferred using a holistic scale justifying their choice by stating that a holistic one is easier to use, is faster and is easier to memorize when assessing. Participants that chose a combination of both considered that using an analytic and holistic tool would allow them to decide when to use which tool depending on the number of students assessed, and the amount of time available for assessment. An additional reason provided was the purpose of assessment, mentioning that when assessment was formative an analytic rubric could be considered while for summative purposes a holistic one would be used. TP32 stated `because you can apply it in several stages of the development of the text. In the process and final product of the text´. Finally, participants who chose an analytic scoring tool considered that these served better to provide detailed feedback to students in terms of areas of the text that could be improved by providing more details and being more efficient for this purpose. TP05 agree to this by mentioning `the grade provided to the student is fairer´.

5.2.3 Participants´ Performance in the Study

In relation to the effects of the training to the participants´ performance in this research study (Table 8), the majority of the TPs (70.8%) mentioned the rubrics provided by the researcher could be useful for future assessment, 20.8% chose to agree with this idea, 4.2%

neither agreed nor disagreed, 2.1% disagreed and 2.1% strongly disagreed. A large proportion of TPs perceived their assessment of the five written samples was easier after the training (64.6%) while 68.8% perceived it more efficient (strongly agreed), 22.9% agreed to the idea of assessment being easier, 4.2% neither agreed or disagreed and 2.1% totally disagree to this statement. A minority of TPs, 20.8% chose `agreed´ when elicited about the efficiency of their assessment post to training, 6.3% neither agreed nor disagreed, 2.1% disagreed and 2.1% strongly disagreed.

The two semi structured interviews conducted elicited the points of view of the eleven participants of Phases Four and Five regarding different aspects of writing assessment and how they changed post to experiencing assessment training. Impact on four different areas of perception post to training were identified: 1) Writing Assessment Procedures, 2) Writing Assessment Scoring Tools, 3) Writing Assessment Training and 4) Teachers as Writers and EFL Assessors. Within each major theme, two to five subthemes of perceptions emerged. The main themes and emerging types of impact on perceptions may be visually depicted on Table 9. The following section focuses on the description of TPs´ responses to the semi structured interviews which lead to the identification of their perceptions and types of impact

5.2.4 Writing Assessment Procedures

As can be noted in Table 9, two different types of impact (subthemes) were identified within this first main theme (Writing assessment procedures): a) an increase of writing assessment importance and b) an increase of perceived importance of student involvement in assessment.

a) Increase of Writing Assessment Importance. TP16 explained that his preference for portfolio use in his classroom as an assessment tool increased. It was mentioned that he now considered increasing the amount of points given to students' writing: dedicating 60% of the grade to portfolio work rather than 50%. He described 'I consider that the portfolio is a better way to evaluate students, maybe at the end of a certain level'. It can be concluded that the main change in perception was an increased level of importance provided to the classroom writing of students and a change in assessment technique preference.

TP325 also reported an increase to the perceived importance of writing and its assessment (first subtheme). She was teaching EFL at a public university at the time of the study and initially (during Interview One) perceived writing as too difficult to teach and to assess in the classroom. However, she reported that she now understood the need for students to develop their writing skills. She explained that she now has more interest in the skill and its assessment and would like to cause the same effect in her students. It was pointed out that students did not want to write and perceived it as too difficult as a direct result of this TP´s perception of the skill. She states so in the following excerpt,

> …there is more interest on my behalf in the sense of not denying its importance in my class…since we did not have enough time and most of them don`t have the interest because they feel English is not for them (the students) …so my task is to change that perspective…showing it as an easy skill and by changing my perspective first.

A second TP reported an increase of the importance given to writing and its assessment was TP23. Her perspective changed once the training sessions finalized, understanding

how developing and assessing writing is important to all language students as well as for all EFL teachers. She mentioned that understanding the importance of writing allowed her to identify the deficiencies that she needed to work on to be able to teach and assess writing. She considered herself as a teacher that still needed much more work to improve her skills and explained,

> …Yes I´ve become more conscious that it is an ability that needs to be taught and assessed. But as a teacher …I am not good at writing so to be able to teach and assess you need to know…

b) An Increase of Perceived Importance of Student Involvement in Assessment.

Regarding her assessment practice, TP62 explained that she considered her students had benefitted from her changes in her assessment procedures. She perceived a positive environment and that they now inquired about their work and about the scoring criteria. For this participant, it had been beneficial because she had been able to reflect on what she was doing, if her students were fulfilling the objectives she had established and above all if they were reflecting on their performance through the assessment of their texts. Therefore, she considered that assessment training was beneficial for students and teachers.

TP313 pointed out that he used to perceive writing assessment as a process that had minimum importance in his classroom. A lack of time and content overload in his program did not allow him to develop it thoroughly. It was stated that students' assessment was considered as one more activity that needed to be done without giving much thought to its actual process. During interview 2, this TP considered that he now gave more thought to the assessment of the skill by acknowledging students´ role in the assessment process and

allowing them to assess their own work. He explained that he now considered possible ways of including students in future classroom assessment activities.

Participants were also elicited their perceptions in regard to the use of scoring tools such as rubrics to assess writing in the EFL classroom. These are more broadly described in the following section.

5.2.5 Writing Assessment Scoring tools

As can be seen on Table 9, within this second main theme different subthemes emerged. For instance, a) teachers´ scoring tool use was the first to stand out. Within this, several subcategories were found such as different positive and negative perceptions in regard to assessment tools. The positive views included those that considered scoring tools useful, easy and/or comfortable to use. However, other TPs had a more negative view and considered them frustrating for teachers and students, or too difficult to use. Subsequent subthemes that emerged were b) analytical scoring tool preference, c) writing checklist preference, d) a change of preference of assessment tools and finally e) an increase of perceived importance of the use of assessment tools. The following paragraphs describe TPs´ semi-structured interview responses which led to the categorization of their perceptions in regard to the use of scoring tools.

a) Teachers´ Scoring Tool Use: Positive Perceptions. TP64 explained that the use of analytic scoring tools, post to experiencing assessment training, facilitated her assessment practice because their use allowed her to focus her assessment and avoid losing time. She specified,

> I felt it easier when you have an instrument. When you don`t have one you lose time thinking about what the student did or didn´t do. When you have the tool to assess you focus …On the Internet there are a lot of rubrics available but they are not really focused to our reality. So, I definitely need to adapt the tool.

As previously stated, she considered rubrics needed to be context specific so they can reflect the assessment needs of students and teachers. However, this is not always the case and therefore adaptation is needed. She commented that she felt the rubrics provided by the researcher during the data collection phases were too advanced for her students therefore adaptation was needed.

b) Teachers´ Scoring Tool Use: Negative Perceptions. TP37 was the first to comment negatively on scoring tools and her perception in regard to their use. She explained that they made students and teachers feel `frustrated´ especially if teachers did not know how to implement them in their classrooms and students did not have the necessary language proficiency. She considered students needed to be taught how to use rubrics and she did not feel ready to do so as indicated in the excerpt below,

> …I feel that I would be frustrating them, I would feel frustrated. They are full of work…I need to train them and little by little let go of them so they can understand them. I still need lots of time.

c) Analytic Assessment Preference. TP16, a fluent speaker of English, Spanish and German, explained that he considered analytic assessment of writing was more useful because it allowed him to be more precise in students' weaknesses and strengths. He

specified that he tended to use the same scoring tool in his German class because he felt comfortable with it, therefore suggesting his preference for this tool. He explained,

> I use for example, the English rubrics with different aspects also for my German exams. Because I like them. Because you don't evaluate only one aspect. Because sometimes students may be very good at grammar or reorganization of the text.

d) Checklist Preference. Other participants such as TP23 and TP32 added they preferred using a checklist to assess their students´ scoring rather than an analytic or holistic rubric. They explained that they considered their use of an assessment checklist was easier and practical to use than any type of assessment. This can be portrayed in the following extract pointed out by TP23,

> …analytic is just too complicated…I prefer using a checklist in which I can just tick what students have done and what they haven´t…it seems students sometimes do not understand the rubric when I used them last semester…with the list it is easier.

Since training provided triggered participants distinct perceptions, the following section focuses on these and their descriptions.

5.2.6 Writing Assessment Training

In regard to TPs´ perceptions of the writing assessment training provided, the following main themes emerged a) assessment training perceptions, b) ideal assessment training traits, and c) perceptions of assessment training dependency. As portrayed on Table 9, teacher participants perceived the training sessions were practical, useful, beneficial for their practice, and supportive for their practice. In regard to ideal training traits, teachers considered it should be permanent, constant, and obligatory, include a lot of practice, include tool adaptation practice, and tool creation practice. Finally, teachers considered

that the possible impact of training sessions depended on many factors such as a) teacher motivation, b) teaching style and c) teaching personality. These were considered subcategories that emerged as part of the subtheme that focused on training impact dependency. Table 9 further identifies TPs and their specific perceptions while the following sections detail participants´ responses that depicted the views of training impact classification.

a) Perceptions of Assessment Training. In this case, TP325 considered that assessment training can help homogenize assessment criteria and procedures. She explained that her school did not follow a specific structure when it came to assessing students´ skills thus many students were not objectively assessed in their courses. She stated that if all the teachers assisted to workshops, assessment could be standardized in benefit of the students.

On the other hand, TP22 suggested the training could be more effective by adding more sessions that are oriented towards specific stages of the process of writing. In other words, assessment training could have benefitted from additional sessions that link the teaching of writing as a process and the assessment of writing.

TP62 explained that she considered assessment training necessary to understand the importance of unifying assessment procedures and standardizing criteria. This with the purpose of guiding the teacher and allowing positive development in the student. She explained that students may have a hard time adjusting to the new teacher at the beginning

of each term and therefore slowing down skill development if procedures are not standardized. In this regard she specified,

> …each semester we are moved from one group to the other and one gets used to working in a specific way and thinks that students are going to accept it…if I give them a specific scoring tool I get them used to assessing a specific way, then they change their teacher and they may be affected because they will work with a different style or a different scoring tool and it will be hard because they may not understand. So, it is good to take these sessions so that we can be homogenous, so we work at the same rhythm and so students adjust to a single tool in the school.

She stated that the assessment training had been beneficial and useful for her in her context. But it may not always be the case for all teachers.

Similar to her co-worker's opinion, TP73, considered that training is beneficial for those that are experienced teachers because it allows them the opportunity to update their assessment skills especially when routine and habits have taken over teachers´ practice. She explained, `…In my case I feel the sessions were very useful because sometimes we believe that we don`t have time to assist or we forget what we learned years ago…´. Thus, confirming that training is perceived as a process that allows teachers to learn new assessment knowledge but also one that allows experienced teachers to reflect on knowledge already acquired.

In accordance to TP73, TP313 perceived the training sessions provided as useful especially to those who are experienced in the field. He pointed out that at times teachers get involved with their teaching routines and assess students' skills without much thought. Thus, he

considered that training can allow them to update their skills to assess writing and other skills as best as possible.

TP315 pointed out that more time was needed to allow teachers to actually reflect on what was reviewed in the sessions. She considered that two sessions were not enough to help teachers to reflect on its content. Therefore, suggesting that training sessions needed to be provided for a longer period of time and on different occasions.

b) Assessment Training Impact Dependency. TP325 pointed out that the benefit that training could bring to a teacher was subjective because it did not only depend on the training provided. She perceived that other aspects such as teachers´ interest to improve and commitment with their teaching practice also influenced on the impact. This can be depicted in the following excerpt.

> There may be changes if the teacher really is committed to his practice because if I have interest in improving and growing I'll find the tools that allow me to do it and use them in my benefit…it is not so much that training changes us it is not the magic wand that changes the situation but it is the teacher who is motivated and committed to his practice…

A second TP to express her feelings in regard the training sessions was TP23. This female teacher pointed out that the benefit that sessions can bring to teachers depends on their willingness and openness to change. She perceived that changing teaching and assessment methods in a classroom would be difficult if the teacher does not accept that changes can be positive. It is my personal belief that the perceptions that this TP had previous to experiencing assessment training were slightly modified in favour of teaching and assessing writing. Although this change of perception has not been reflected in her

reported classroom assessment, it has been in her own conception as a teacher/writer and in the increase of writing activities implemented in her classroom after experiencing training. TP22 agreed with this view and added that many factors are part of this process such as teaching styles, teacher motivation, and teacher personality among others. He specified that he perceived training as a trigger of teacher reflection that allowed him to analyse how he handled teaching and assessment in his classroom as he states in the excerpt below,

> …I think it depends on the teacher… It is difficult to boil teaching to a single style. In my experience having training does not change my teaching practice but definitely puts things in perspective and it helps me think about my own practice …because I try to do it myself as best as I can… and having another perspective into my teaching style and having myself think over things that I do and if I do them well and the ideas that I am being taught or shown really work or not helps me think about the practice and improve its quality. I think that is my personal case…

Similarly, TP62 perceived that impact that training may or not depend on the teacher´s personality and personal motivation. She perceived that in her case it had been positive because, very similarly to TP73, it had provided her with the opportunity to take time to reflect on her assessment practice. She noted,

> I think it depends on the teacher if you really do your job and are aware that students depend on you and that at some point they need to pass an exam then training may have an impact but if you come to work just for the sake of it and only for the money and you do not give much importance to your job then the impact will be negative… if people do not have the willingness to learn and to change attitudes and habits nothing is going to make them change…In my case it was positive I began to think a lot about what I am doing, if what I am doing is right or wrong. We stopped to reflect.

TP73 considered that the benefit of training could be increased if more people from different schools were present in the session therefore allowing for the difference of opinions to nourish the sessions. She added that when institutions obligate their teachers to assist to training sessions, the benefit of these may be hindered. However, she specified that for her, and for other teachers, experiencing sessions and the possible improvement their teaching and assessment practice could have was her contribution as a professional. She also considered it a way of defending it among those who do not take teaching of EFL seriously. The following excerpt portrays these perceptions.

> I believe that there can be change…maybe most of the teachers assist to the workshops obligatory but if you assist with your mind open, consciously open to what you are listening to and experiencing can actually happen then change may be possible…but it is very important to help our area and contribute…sometimes there has been a lack of support to the field but if we don`t take our profession seriously then no one will.

c) Ideal Assessment Training Traits. Once more TP325 considered the sessions had been beneficial for her because she had seen ways of improving her practice. However, she recommended that sessions be constant throughout a term and more frequent so that the teacher is provided a follow-up that can link practice in the classroom with theoretical support.

TP23 suggested that sessions be provided more frequently during the term, more context specifically and more practical. She added that although the training sessions encouraged assessment practice during the workshop she felt more needed to be done to clarify teachers´ doubts. She explained

> …more frequent and more practical to learn how to actually create rubrics and not only to use them. Lots of practice so we can learn…sometimes the textbook includes rubrics but it is better to create or adapt one that can suit your context and your needs.

Very similar to TP23, TP32 commented training needed to be permanently implemented to cause effect or to help those that do not have an academic background. It was also pointed out that teachers need to understand how to construct a rubric. Therefore, training sessions also need to be oriented towards this purpose so that classroom teachers know how to construct an assessment tool. In this sense, he commented,

> …we learned the language in English courses or in the states (USA), nobody taught us how to teach the language. We do not know if it is correct or…how it is applied in the classroom. We have never taken a training session or a course to teach and I believe they should be permanent.

Finally, he explained that assessment training motivated him to reflect on his practice and his weaknesses as an EFL teacher so they could be improved.

TP37, considered that training needed specific characteristics for it to be of assistance. She added that she considered a teacher improved practice if training and academic professional development is constant and over time.

Finally, TP315 considered that the only way of actually generating change was forcing teachers to attend training sessions through the school´s administration. It can be concluded that the inexperience that this TP had at the moment of the study and her need to

be guided in her activities may have influenced this perception of the need of obligation with assessment training.

5.2.7 Teachers as writers and EFL writing assessors

The fourth major theme that emerged, and as is visualized in Table 9, was teachers´ views of themselves as writers and as assessors of EFL writing. Subthemes that emerged focused on participants´ positive and negative perceptions about themselves as writing assessors and writers. On the positive side, participants found their participation in assessment training improved their practice, allowed them to feel more supported, more confident, more motivated, more focused, more careful and less afraid when scoring writing (emerging subcategories). On the other hand, there were those that felt bad, disappointed and unskilled. In this sense, the following paragraphs further explain these subthemes and subcategories that emerged from transcript analysis. Table 9, included below graphically portrays teachers´ reported perceptions.

a) Teachers as EFL Assessors of Writing: Positive Views. Firstly, TP16 indicated that experiencing assessment training allowed him to feel supported and confident about implementing change in his assessment methods. He perceived that changing long-time established assessment methods was a good idea that could actually benefit students as he specifies by explaining `I think when I was in this training…this training session … it gave me also ideas and also the support to believe that I am not the only one with this idea…´ Therefore, explaining he felt more confident when sharing his experiences during the sessions.

TP313 explained he had previously perceived himself as a teacher who had deficiencies as an assessor. However, post to training he explained that he felt he had improved as a teacher by providing students with that extra help that they need to develop writing. He stated,

> I feel I am doing a better job because I used to only revise the vocabulary and grammar use, which I believe is not assessing writing, it's only compelling with the evaluation without really paying attention at how you are doing as a teacher and without helping the student. I feel I am a better teacher because I am helping them improve

Regarding her perception of herself, TP315 pointed out how training has allowed her to change her perception of herself as a BA student and as a teacher. As a student, she explained that she now valued more the assessment of her teachers and the feedback she received. As a teacher, it also changed her perspective of what `should´ be assessed in students´ language proficiency and what `should not´. She added that she now felt more confident and secure of what she was doing as pointed out in the following excerpt,

> …personally, as a teacher and as a student I have noticed…a way to balance things…I cannot assess only for the sake of it…so yes it totally changed the way I see things…I don't feel "dirty" anymore every time I score a text…I have the base established of why you obtained a 9…because of this and this (signalling the rubric) here it is, here is what I did…so I am not afraid anymore…more confident.

b) Teachers as Writers and EFL Assessors of Writing: Negative Views. The second TP to comment on this was TP32. He considered that training had allowed him to change his perception of himself as a teacher-assessor. He specified that he felt disappointed with his practice because he noticed he needed to improve to become a good teacher and a good

assessor. However, he perceived himself as more motivated to look for more training sessions that can help him improve more permanently.

These views were shared by TP37 who considered that she needed to work on her assessing and writing skills. Like TP23 and TP32 this teacher considered she had been able to analyse her skills as a writer and concluded that she needed more work to improve her writing abilities. It was pointed out that she believed that teachers need to be writers in order for them to teach writing so that they could feel secure enough to teach and assess the skill. She explained this by stating `We also need writer training; we need to feel more confident as writers. We need more fluency when we write´.

Table 9 Teachers´ Perceptions of EFL Writing Assessment: Interviews

Category	Type of Impact	Teacher Participant
Writing Assessment Procedures	More importance to Writing Assessment	TP16, TP23, TP32, TP313, TP325
	More Importance to Student Involvement	TP313
Writing Assessment Scoring Tools	More Importance to Scoring Tool Use	TP23, TP32, TP62, TP73
	Change of Scoring Tool Preference	TP22, TP313
	Analytic Scoring Tool Preference	TP16
	Checklist Preference	TP32, TP23
	Scoring Tool Use	Useful: TP16 Comfortable to Use: TP16 Easy to Use: TP62, TP22. Frustrating to use: TP37 Difficult to Use: TP23
Writing Assessment Training	Writing Assessment Training Perceptions	Practical: TP23, TP325 Useful: TP32, TP313 Beneficial for practice: TP62, TP73, TP325 Supportive for practice: TP16, TP32
	Ideal Assessment Training Traits	Permanent: TP32 Constant: TP37 Obligatory: TP325 Include practice: TP62, TP325 Include scoring tool adaptation: TP23, TP32, TP37 Include tool creation: TP23, TP32, TP37
	Assessment Training Impact Dependency	Teacher Motivation: TP22, TP23, TP62, TP325 Teaching Style: TP22, TP23, TP62, TP325 Teacher Personality: TP22, TP23, TP62, TP325
Teachers as Writers and EFL Writing Assessors	Teachers as EFL Writing Assessors	Improved Practice: TP313 Felt supported: TP16 More confident: TP16, TP62, TP315 More motivated: TP32 More focused: TP32 More careful: TP313 Felt Bad: TP313, TP325, TP32 Disappointment: TP313, TP325, TP32
	Teachers as Writers	Felt unskilled writer: TP313, TP325, TP23, TP32

5.2.8 Section Conclusion

In conclusion and to answer RQ2 (*What is the impact of assessment training on teachers´ perceptions of writing and on their perceptions of classroom writing assessment?)* it can be stated that the subcategory that involved the most participants was `more importance

teachers gave to the skill of writing and its assessment in the classroom´ (five TPs) while the subcategory that had the least participants involved was `analytic scoring tool preference´ (one TP). This would suggest that assessment training may be used not only to bring improvement to the assessment process in the classroom but also to raise awareness of the importance of assessing writing abilities. Finally, more TPs had positive perceptions in regard to ideal training trait and impact dependency while more TPs had positive perceptions of themselves as assessors and their experience as assessors after training sessions.

On the other hand, answers provided to the post training- online questionnaire by the 48 participants of Phases One, Two and Three confirm that writing assessment training encourages teachers to give more importance to teaching writing and its assessment in their classroom. It also can be concluded that their use of scoring tools such as rubrics was considered more useful, practical and easy after receiving assessment training. Table 8 depicts the results obtained from the responses of the 48 participants to the online post-training questionnaire.

This project considered that by understanding all of the participating stakeholders´ perceptions of writing assessment training, the assessment procedure each institution follows may be further understood and improved. Therefore, Section 5.3, included below, describes the findings related to the perceptions of a second type of stakeholder: the EFL programme managers of the participating institutions.

5.3 EFL Program Managers´ Perceptions of Writing Assessment

The third research question of this study (RQ3 *What is the impact of assessment training on language program managers´ perceptions of writing assessment?)* focuses on the results of the analysis of the transcriptions of the eight interviews conducted with four EFL Language Program Managers (PM). Three main emerging themes were identified 1) Perceptions of the Nature of Writing Assessment in the Mexican Classroom, 2) Perceptions of the Importance of Assessment Training for EFL Teachers and 3) Perceptions of Impact of Assessment Training. For instance, in regard to this first main theme, all the EFL PMs agreed that writing is an important skill that needs to be taught and assessed in the classroom.

Additionally, PMs added that the lack of time to include the teaching and assessment of writing in the EFL language curriculum and the EFL classroom is the biggest constraint. All the TPs stated to include writing in their teaching and assessment practice giving it different degrees of importance. Only one, participant stated to not teach and assess writing because it was not part of their overall learning goals being time the biggest constraint. The perception of only one PM converges with these teaching practices. PM1 stated that writing was a secondary aim in their program since they required their students to obtain 500 points in the TOEFL ITP as a graduation requirement. However, the other PMs identified other difficulties that were not included in TPs answers. Difficulties such as teacher training, teachers´ writing background and students´ low writing proficiency were among the identified constraints.

Table 10 Language Managers´ Perceptions of Writing Assessment and Assessment Training

	Writing Assessment in the Mexican EFL Classroom
PM1	Writing was not a priority therefore not assessed, post to training more attention provided to it. Lack of time biggest issue, overload in EFL program
PM2	Focus of assessment was skills for TOEFL ITP. Teachers do not write so they do not teach it. Time issues and lack of writing skills on behalf of the teacher and students.
PM 3	Writing Assessment depends on the teacher and her assessment style. Issues were lack of time, lack of teacher training. Tests, standards and rubrics provided by administration.
PM 4	Writing Activities focused on grammar accuracy, it is limited by the amount time available and students' poor writing skills.
	Importance of Assessment Training for EFL Teachers
PM1	Important to implement change and improvement in teachers and decision makers in a program. Training needed, knowledge constantly being updated.
PM2	Training is important to update experienced teachers' assessment and teaching practices and encourages motivation in their work. It is needed to focus on developing writing in the teacher.
PM3	Training provides opportunities to change teachers' perspectives of writing. Training is necessary to understand how to teach writing; and unify assessment standards.
PM4	Training gives updating to experienced teachers.
	Impact of Writing Assessment Training
PM1	Change in EFL program's goals, inclusion of writing in the program. More activities were incorporated to the daily activities.
PM2	Changes in perceptions of writing, consider other aspects than language accuracy when assessing. Reflect on the need of the institution to provide development to teachers.
PM3	Redesign of writing assessment tasks, emphasize direct writing assessment, planning of future training sessions, Possible change of assessment activities, learned to value teaching staff.
PM4	Considered the institution should provide training to teachers, and provide enough time and training sessions for teachers to improve.

In regard to the impact of assessment training on language managers´ perceptions, it can be argued that PM1 and PM3 reported to have perceived writing assessment differently post training. Although both of them prior to training considered writing an important skill to develop they did not give it this importance in practice. PM1 explained that she now was a

witness to how teachers were more active in including more writing activities in their EFL lessons this leading to her initiative of bringing forward to other school authorities the importance of encouraging writing in the regular assessment of the teachers. PM3 on the other hand, explained that assessment training had allowed her to actually reflect on what was being done in the institution and the tools that were being used to assess the language. She pointed out that the writing tasks that were included in the unit exams were updated. Finally, the teacher indicated that she had initiated the planning phase of changing the assessment procedures and assessment tools of the institution. PM2 and PM4 did not specify a specific change post to having experienced assessment training.

The following section has the purpose of providing a wider perspective of each PM and the provided answers that lead to the emerging themes previously described and portrayed in Table 10.

5.3.1 Writing Assessment in the Mexican EFL Classroom

PM1 considered that the EFL program at the university intended to provide students with extra resources to support their professional teaching practice. Students were, at the time of the study, enrolled in a BA program that prepared teachers to teach in elementary education. Therefore, she considered that a deep preparation in the use of language skills was not necessary. She explained that the program intended to develop the four language skills being writing one of them and teachers needed to adapt their contents to be able to fully exploit writing in the classroom. She added that the major constraint to the assessment of writing was the lack of time therefore resulting in teachers´ and students´ rejection of the skill. She explains,

> …for the teacher, it is an issue of time because in theory we have 4 hours a week however the work dynamics here is quite peculiar. Students have periods of practice (internships) in the elementary schools… they have lots of work and need to prepare lots of material…and that affects directly our practice because we don´t see them during that time.

In relation to the assessment of writing, PM1 explained that she perceived that writing was not given importance when assessed because it was not students´ priority and added `…for many of our students here …it is not their priority and some don't even like it´. Thus, suggesting that English was an additional class that did not have importance for students. On the teachers´ side, she added that they preferred to work with the little time they had with skills that students could manage to work with such as grammar, vocabulary, speaking or reading thus leaving aside writing due to its difficulty.

The second language program manager (PM2) mentioned that since their overall goal was to lead their students to take the Institutional TOEFL test, their focus in the program was listening, grammar and reading. Although writing was given treatment in the classroom she considered teachers gave more importance to other things. This stakeholder stated,

> …In fact, writing is part of the program, it´s included but what happens is that some teachers do not give much importance in the classroom…because everyone has freedom in the classroom to focus on what they consider important. And the program is quite heavy...We would like to have time for everything… We have 75 hrs per term…so it´s never enough…. but I think writing depends on the teacher, and the style of each teacher.

When asked about the issues of considering writing in the EFL program, this PM considered that time and the lack of writing on behalf of the teacher made the development of writing more difficult in the classroom. She pointed out,

> One of the things that limits is the time and because the teachers do not write. If your teacher does not write that is if nobody ever taught them to write then they will not teach it...and time ... if I put something in writing, is something I need to return to see how they improved …and that takes up a lot of class time...

The third language program manager (PM3), reported writing had an important place in the language program at the institution. She believed they still needed to encourage its assessment among teachers. She pointed out that although the administration provides the monthly exams to teachers, portfolio guidelines and assessment rubrics, assessment is subject to the teacher and to their own view as can be shown in the following excerpt,

> …assessment has a lot to do with the teacher's personality, but I think it is something that the teacher does daily and enthusiastically but he lacks instruction. It is one of the points, and also the lack of time to assess…

PM3 considered that one important issue was the lack of teacher training in the teaching and assessment of EFL writing and a lack of teachers that write. In an additional comment, PM3 added that another difficulty faced in language programs is students´ poor writing skills. It was also explained that many of the times students did not know how to write in Spanish making their writing in English more difficult therefore `the teacher loses a lot of time explaining things that students do not know about writing´.

The final language manager, PM4, an English coordinator at Institution B, considered writing was part of the EFL program, but explained that she and other teachers considered grammar lesson notes as students´ writing of each unit and assessed it as a monthly portfolio. She specified that for her and the EFL program, writing mainly focused on grammar accuracy. She stated `I evaluate writing based on a portfolio, I give all the

structures, and they write with me. I write on the blackboard and they write along with me´. PM4 adds to this comment that `students have poor writing skills, they do not manage to connect their ideas´. This participant considered that constraints to the inclusion of writing in the program were in their majority teacher/student issues being time the biggest constraint and students´ lack of reading skills.

5.3.2 Importance of Assessment of Training for EFL Teachers

In relation to assessment training PM1 considered that training is necessary to conduct their job and added that knowledge is constantly being created and teachers need constant updating. She commented,

> … It´s necessary ... continuous professionalization because there will always be new techniques There will always be something new ...a BA, a master´s degree, even a doctorate you learn what is known until the moment... and then what? Life does not stop there and knowledge either…

Regarding assessment training, PM2 considered that it is 'very necessary' because teachers would not have the sufficient knowledge to teach writing. Additionally, she believed training could not only focus on assessment but also on how to write. In other words, training should be a sequence of sessions that include how to become a writer, how to teach it and how to assess writing. In this sense, the PM2 pointed out

> ...it is quite necessary. Because if you do not write as a teacher, how will you teach it if you do not know. It is very necessary that teachers practice it and experience it... Should have training… first of how you will teach it and then complement it with the evaluation...Here for example writing is part of the evaluation…

She continued explaining that she believed teachers could benefit from training sessions and if there was sufficient time, then it may give them opportunity to reflect on what they are doing in their classrooms. In relation to the need of training, as with PM2, this administrator considered that providing teachers with training is very important because it allows them to recall what was learned as a novice teacher

> ...I think it's very important, and it is very important to be an updated teacher because I've seen teachers who are in the ICELT (In Service Certificate of English Language Teaching) ... and they wake up. It's like a brainstorming session, that gives you motivation, and then it goes to waste or decreases and we must again refresh everything. Apart from training from the start, it refreshes teachers on what they already know...

5.3.3 Impact of Writing Assessment Training

PM1 considered that the training sessions provided could help teacher participants improve their inclusion of writing in their activities. In other words, and according to her perception, teachers would not have enough time to improve the assessment of writing but instead would only begin improving their teaching of the skill.

In Interview 2 (post to training), she stated that she had noticed teachers included writing in their everyday informal discussions in which teachers shared their experiences and opinions about writing in their classrooms therefore perceiving teachers increased the importance they gave to writing in the classroom. This PM stated that although the institution did not allow a change to the curriculum, she considered that after experiencing training sessions, the way writing is perceived and taught in the classroom could change. She stated,

> … I have seen a change because I have never heard them talk about skills or how they would assess... I knew teachers considered implicitly writing in their lessons... I have heard them talking about their corrections in the writings and ... now they are giving more importance to the topic (writing assessment) ... We are not autonomous so I can´t implement a change to the curriculum…but a change within the way we teach our classes ... writing was inside and was not a trait to be evaluated as such ... at least we could try ...the training came to open our eyes and helped us remember the importance that writing has…

With this statement, PM1 considered that a change that could be worth trying is considering writing as part of the skills to be evaluated more formally within the classroom. This could also raise awareness of the authorities that may notice the importance of the inclusion of writing in EFL programs.

For PM2, assessment training allowed her to realize that training needs to be implemented one step at a time. Therefore, it can be concluded that for this PM training sessions allowed her to modify her perception of the importance of providing teachers with opportunities of developing professionally.

During Interview 2, PM3 mentioned that she personally felt how she had started recalling information she had reviewed when she was a young and novice teacher and continued explaining how as a manager she had initiated planning the implementation of changes in the assessment tasks and criteria the language centre had established. These ideas can be portrayed in the following excerpt,

> It helped me...often you know the things but you focus on the easy things... It refreshes knowledge. But I think it helped, it helped us change a few tasks of the exams. For example, some activities were changed such as a recipe to something that involved more development or more communication... Mainly it changed the perspective of what is writing, because sometimes we focus only on grammar, but instead look at it from a whole where the student orders ideas and communicates ideas.

She added that she noticed some teachers were more worried about writing and its assessment, especially in the low levels because many of the times they focused only on grammar or spelling, but post to training they paid more attention to the actual meaning that students want to communicate. It was indicated that the coordination was analysing the possibility of using only portfolios to assess writing because after the training session, teachers reflected on their actual purpose of their assessment and had communicated to the office that the test did not actually tell them what a student learned. Another possibility that was mentioned was the implementation of a writing course in which writing skills could be practiced with teachers and students.

Finally, PM3 added that she perceived the training sessions had a positive impact because she had noticed teachers at this centre were very good teachers committed to their work and eager to be updated. She added that constant training is needed to avoid having teachers get involved deeply in their routines or heavy workloads. Additionally, she mentioned that her biggest constraint faced was time and added,

> I think that the problem with writing assessment is that it requires extra class time. If we consider teachers have to find another job to pay their expenses so they are overworked and do not have the extra time to devote to the evaluation or get to work with students…and not even to training.

In relation to assessment training, the PM4 reported that she did not have enough time to implement change or analyse what could be implemented to the classroom.

5.3.4 Section Conclusion

In conclusion and to answer RQ3 *(What is the impact of assessment training on language program managers´ perceptions of writing assessment?)* all the EFL program managers agreed that writing is an important skill that needs to be taught and assessed in the classroom. Additionally, PMs also agreed that the lack of time to include the teaching and assessment of writing in the EFL language curriculum and the language classroom is the biggest constraint. Only one of the PM's answers was in alignment with TPs assessment experiences.

PM1 stated that writing was a secondary aim in their program since they required their students to obtain 500 points in the TOEFL ITP, therefore TPs did not give it much time in the classroom. However, the other PMs identified other difficulties that were not included in TPs responses. Difficulties such as teacher training, teachers´ writing background and students´ low writing proficiency were among the identified constraints.

In terms of the impact of assessment training on language managers´ perceptions it may be concluded that PM1 and PM3 reported to have perceived writing assessment differently post to training. Although both of them prior to training considered writing an important

skill to develop they did not give it this importance in practice. PM1 explained that she now was witness to how teachers were more active in including more writing activities in their EFL lessons, which lead to her initiative of bringing forward to other school authorities the importance of encouraging writing in the regular assessment of the teachers. PM3 on the other hand, reported that assessment training had allowed her to actually reflect on what was being done in the institution and the tools that were being used to assess the language. She pointed out that the writing tasks that were included in the unit exams were updated. Finally, the teacher indicated that she had initiated the planning phase of changing the assessment procedures and assessment tools of the institution. PM2 and PM4 did not specify a change post to having experienced assessment training.

As previously mentioned, this project considers that by understanding the views of different stakeholders in the language assessment process, classroom assessment may have more possibilities of being conducted under standardized procedures. Therefore, this project also considered the students´ views of the assessment of writing and teacher training.

5.4 Students´ Perceptions of Writing Assessment and Writing Assessment Training

The fourth research question *(RQ4 What are students' perceptions of EFL teachers´ regular classroom writing assessment and of the importance of writing assessment training?)* focuses on the perceptions of a different stakeholder of the assessment process: EFL students. As depicted on Table 11, data was obtained from two focus group interview sessions conducted with four groups of students (one from each of the four institutions

under analysis), adding up to eight focus group sessions. Results reveal that students´ perceptions focused on four main aspects of writing assessment: a) notions of the nature of writing assessment, b) the current classroom assessment of writing c) their points of view of teachers´ current classroom assessment and d) their views of the importance of teacher assessment training.

In regard to main theme A, the nature of writing assessment, students from all the institutions considered that assessment should be accompanied with specific feedback that can help students identify the areas that need to be improved in their texts. Students from Institution A pointed out that writing assessment should be balanced among the rest of the skills in the classroom while all the students considered that writing should be a priority of all the school and all English teachers.

In regard to main theme B, the reality of writing assessment in the EFL classroom, students from Institution A and B reported that they considered the importance given to the teaching and assessment of writing was not enough considering it to be `superficial´. It was explained that in reality teachers did not actually assess writing based on their performance. Instead for teachers it was enough for the task to be handed in. Thus, students from Institution A and B disagreed with this assessment practice considering that a teacher should focus on what was actually done. It was reported that teachers tended to omit explaining the criteria they would consider to assess writing therefore students did not know why a specific grade was given to them. They considered this practice could impede their development instead of helping it. In this regard, students considered feedback was crucial for them to improve their written skills.

Regarding main theme C (students´ views of teachers´ current assessment practices), most of the student participants expressed to feel positively about teachers' assessment methods while a few expressed negative perceptions. Students of Institution B pointed out during both focus group sessions that they felt `uneasy´ with the teacher´s assessment methods because they were not informed of the criteria neither given the opportunity to improve their work. Additionally, it was reported that writing was not considered in the monthly assessment of the class. Participants of Institutions A, C and D mainly reported positive perceptions of teachers´ assessment of writing (felt supported in their language development, felt given importance because their work is read, fearless to be assessed, and comfortable with assessment procedures). However, students of Institution D explained that even though teachers made a big effort to make the assessment process a learning experience, they found the assessment tools used by the institution were `too heavy´ and `too difficult´.

Table 11 Students´ Perceptions of Classroom Assessment and Assessment Training

Focus Group	Nature of Writing Assessment	Current Classroom Assessment of Writing	Perceptions of Classroom Assessment of Writing	Importance of Teacher Assessment Training
A	All the participants considered writing assessment of major importance for language development.	- Superficial treatment of writing - Assessment not based on performance. - In favour of the use of technology (Social Media) to assess writing. - In favour of self-assessment.	-Felt supported - They felt important because teacher read their work, - Fearless	Training very important to update assessment skills of experienced teachers.
B			- Felt uneasy - Disagreed with assessment methods.	Training very important accompanied of good rapport with students.
C		- Prefer assessment accompanied of feedback	-Felt comfortable - Assessment on occasions too lenient	Training needed to connect reality with classroom assessment practice.
D		-Sometimes too lenient with assessment. - Peer assessment favoured. -Monthly assessment not accepted.	- Felt comfortable.	Training needed to update skills of experienced teachers.

Finally, of Theme D, which corresponds to students´ perspectives of the importance of teacher assessment training, all the participating students considered that training was important because it is not only enough to have knowledge of the language but it is important to know how to pass on the knowledge of the language. Students from Institution A stated they perceived training was most important for inexperienced teachers so they could update their practice. Additionally, students from Institution B pointed out that the interaction that a teacher can build with a student is more important than any other qualification. Table 11 portrays the information previously described while the following

sections describe in detail the responses provided by the student participants in the focus groups conducted.

5.4.1 The Nature of Writing Assessment

All of the student participants in the focus groups conducted at Institution A considered writing assessment was important for their adequate development of English especially because they considered the four skills (listening, speaking, reading and writing) need to be treated equally in the classroom. It was pointed out that they were in favour of self-assessment with the use of a correction code by stating `…I like it when the teacher writes the symbols on my writing…because I have think about what I wrote and why it is wrong…I need to hand it in again corrected…´. However, it was clarified that feedback was given occasionally depending on the time the teacher and students had available. Students specified the feedback provided was adapted to students' availability and needs exemplifying how electronic media or social networks were used to provide feedback to students.

Student participants from Institution B specified that the inclusion of writing is of major importance to their language development. Students described that they considered an EFL program needs to provide time to their students' learning and explained,

> I believe we still need good teachers…we need time…here are some good teachers and some others that are not…and time…here time is a problem…the previous teacher did not work equally…but personally I feel very comfortable…I feel satisfied.

The student made reference to the limited amount of time that is given to classroom learning and assessing in their EFL program therefore considering time as a crucial factor.

5.4.2 Current Classroom Assessment of Writing

Students from Institution A reported to have different perspectives of how writing was assessed in their institution. However, all of the students agreed that teaching and assessing writing was not a priority of the EFL program they were studying or of their teacher. They perceived writing was taught very superficially and the subject of English in general is given a ` very shallow´ treatment. Students of this institution explained that they would like to see their teachers and school authorities give more importance to the development of the skill in the classroom and its assessment. When interviewed post to training, students reported they now received feedback and were asked to check their work and analyse it with the use of a correction code. They perceived their teacher was more challenging with the texts that they were required to hand in and was more specific with the feedback received. They considered `…it's ok because now I know what is wrong and right´.

Students who participated in the focus groups conducted at Institution B considered writing assessment key for the development of a language. However, they pointed out that in their classroom it did not receive any importance. Every month they took a test, handed in an activity portfolio and did extra homework. Thus, writing was not considered part of the criteria they were evaluated every unit. It was pointed out that the same assessment methods and techniques were used throughout the term, suggesting that assessment

training did not have a significant impact on classroom assessment practice or student perceptions.

A different group of students, those studying at Institution C, explained that previous to the assessment process the teacher explained the criteria that would be considered because topics and tasks were different every unit and therefore assessment criteria was different. Additionally, the proficiency level they were studying also was considered and explained, `...he adjusts to our level…we can´t have everything perfect and he adapts to that when he gives us our feedback…´ Students added that although they felt comfortable with the assessment of their texts, they would like to receive personalized feedback about their texts instead of using symbols to correct work. Students explained that their teacher made a great effort to provide feedback to their texts but they would like to receive a more detailed explanation of the scores obtained. In other words, students perceived the need to elaborate on the scores provided to the text by the teacher.

Students who studied at Institute D considered their teacher gave significant importance to the development of writing in the classroom. But, students clarified that the teacher followed a very specific assessment process in which students were firstly engaged in a peer assessment process and then papers were improved and handed in to the teacher. Student A explained,

> For example, first with my colleague he gives me his writing and we exchange notebooks and he checks my work and I check his and then we give it to the teacher and the teacher now checks if you 're okay.

Students then continued explaining that this peer-assessed work was part of their portfolio activities and represented 50% of their final grade while the monthly test was the other 50%. Throughout the second interview, students reported that now their teacher adapted his assessment criteria according to the purposes of the unit. In each unit, the professor explained to students what the writing task was, what was expected from their performance and the focus of his assessment. Student B explained,

> Change ... I think he observed the group and fit the change to our needs and I think that although we are different ... a group everyone has a style ... the teacher was not always the same ... there were times when he changed his ideas and criteria…

5.4.3 Perceptions of Current Classroom Assessment of Writing

In relation to the assessment of writing, students at Institution A explained that they did not agree with what teachers were assessing. Prior to training assessment, students indicated that their teachers did not assess their work based on their performance, as depicted in the excerpt stated by Student C below.

> …I believe that this is something bad because she is not really evaluating your performance or how you advanced…there could be someone who knows very little or knows a lot…and just handing in something and it doesn't matter if its ok or not and there may be someone who gives very little importance to the work and they will have the same grade only for handing it in…

Students reported they felt they were 'losing time' or the teacher was even 'impeding development' because they were not given the opportunity to improve their work.

When interviewed post to training, students from Institution A pointed out that they felt more supported and secure because they were sure their teacher read their work, which resulted in their increase of the importance given to writing as well. Student D stated

> I am not afraid anymore to write I used to be afraid of making mistakes and now it's different… I have the security that the teacher reads it…not only signs it…and she tells us everything, where we were confused and now it is more important because we didn't use to write…only sentences and now we write texts…

When interviewed prior to training, students from Institution B stated to feel uneasy with the teachers' assessment methods because they were never given any feedback or opportunity to improve. Students perceived they were given an unfair treatment because they were not given an opportunity to improve their work as Student E stated in the following excerpt,

> …for me personally ... for example when evaluated sometimes one comes out very well in the test and at least I do not understand why sometimes one gets very low ratings ... with her about 8 and if one gets 79 you do not know what went wrong with the decimal... like any little thing lowers your grade very much and sometimes I feel that that is the way she scores ... in the portfolio she makes observations but she never tells you if you were wrong or right ... one cannot know the mistakes you made…We know we are not good…but I would like to know what I did wrong.

In relation to the assessment of writing, students studying at Institution C reported they felt `fine´ with the current assessment methods of their teacher and specified that the teacher had assessed their work continuously throughout the term and at times was too lenient with their work as specified below,

> …I feel ok with the grade…I sometimes feel the teacher gave me more than I deserve…if I wanted the 10 I need to write much better …there weren't many changes, she scored our work the same thorough out the term…

Participants from Institution D explained they felt satisfied with the current way the teacher was assessing their writing and stated they considered the institution also gave sufficient importance to the skill. It was explained that the teacher had recently implemented a correction code to facilitate students' self-assessment which students found adequate. Students added that they felt comfortable with the assessment practice of their teacher. However, it was explained that they perceived the strategies to evaluate writing and the test applied every month were 'too heavy' and 'too difficult'. Student E explained,

> …I would have liked to change the test…it was very heavy ... I would have also liked him to be more specific with his feedback and not only tell us what is wrong ... I mean specify if we need to change something...

Therefore, considering it needed to be changed. Additionally, they considered that they would benefit more if the assessment feedback were more specific in relation to what can be done to improve their written texts.

5.4.4 Importance of Teacher Assessment Training

Focusing on the importance of teacher assessment training, students from Institution A clarified that they were not aware of the professional background their teacher had but considered it was essential for them to be professionalized to assess the language by stating ` it is important for the teacher to be prepared because if she isn't then how can she teach us? How will she know how to evaluate us...´ It was also found that one of the most significant changes in students' perceptions post to training was in the amount of importance they gave to writing in their language learning.

In relation to assessment training, students studying at Institution B specified that they considered teachers needed to be professionally prepared and trained. However, they all agreed that it was equally important that teachers have good rapport and interaction with the students as indicated in the following excerpt,

> I cannot remember if she is prepared... it is very important for teachers to be prepared…but also the interaction with the students is very important ... because the fact that you have knowledge does not mean that learning is meaningful ...

From Institution C, students considered that EFL teachers need to have specific teacher preparation and training that can allow them to understand the best methods to assess the language and above all to teach a language. Student F explained,

> It's very important because maybe it depends on that …for example the teacher gives us dynamics…but it is not the same if the teacher knows about language pedagogy…or if she only knows English maybe she's good but the methods of a teacher need to be better…It´s the same when she assesses our work.

In this regard, Student D mentioned that both language skills and teaching skills need to be present in a teacher. Therefore, giving importance to language and teaching competence as depicted in the extract below.

> …They can talk very well but they cannot explain well ... so it´s very important to be trained as a teacher and study that ... I cannot come to teach you something that I did not study ...I would not have the same sense of connection…

Finally, in relation to teachers' assessment training experience, students from Institution D considered experienced and inexperienced teachers should be academically prepared to assess language skills, especially the experienced ones so they can update their skills. They

explained that it is not enough to know English, but it is necessary to have teaching and assessment skills that can only be acquired through professional preparation.

5.4.5 Section Conclusion

As shown on Table 11 and in regard to teachers´ current classroom assessment, all the student participants considered that writing needed to be developed in the classroom for their language skills to be fully developed. It was also found that students from Institutions A, B and C found that peer and self-assessment could be useful to improve their texts as a method to implement before having the teacher assess their work. Learners from Institution C expressed they would prefer having a scored accompanied of specific feedback that could help them understand their performance and how to improve it.

In regard to participant perceptions, students from all the institutions except one perceived their teachers´ assessment to be fair and comfortable to work with. Participants of Institution B felt uneasy and disagreed with their teachers´ assessment techniques stating that other practices would be preferred; while students from Institution C felt comfortable with the assessment received but perceived their instructor too lenient with their scores.

When elicited about the importance of teacher assessment training all the participants agreed it was important but expressed different reasons for its importance. For instance, focus groups conducted in Institutions A and D agreed that training was most useful for those experienced teachers to update their skills. Students in Institution B considered that training was important as long as its accompanied by good rapport skills with students,

stating that training is not useful unless the teacher has good communication skills with students. Participants of Institution C considered assessment training was a tool that should allow teachers to connect the reality with the classrooms.

In an effort to answer RQ4, it may be argued that all the participating students had positive perceptions of assessment writing and of the importance of teacher assessment training while all except one focus group perceived positively their teachers´ assessment methods. However, teachers´ have also been known to influence assessment with personal variables such as gender, teaching experience, academic background among others. The following section points out the results obtained from the analysis of teachers´ personal variables and their role in writing scoring.

5.5 Role of Assessment Training and Teachers´ Personal Background on Analytic and Holistic Assessment of Classroom Writing

As mentioned at the beginning of Chapter Five, RQ5 (*To what extent does writing assessment training and teachers' personal background impact their use of analytic and holistic scoring tools to assess opinion essays in the EFL classroom?*) focuses on the quantitative analysis of instructors´ analytic and holistic scores of five written samples of students´ classroom writing. Additionally, teachers´ distinct personal characteristics such as gender, teacher experience and academic background are also analysed for their possible influence on the scores provided.

Analytic and holistic scores were introduced to the SPSS software program V.23 and descriptive statistics such as Mean, Mode, Frequency, Maximum Score, Minimum Score

and Standard Deviation were run with the purpose of understanding the nature of data. Inferential statistics, Reliability tests (ICC), a Paired Sample T-Test and an Independent Sample T-Test were calculated to understand the consistency of scores among teachers, and how specific teacher characteristics and assessment training had or not an impact on them. The following section focuses on the description of the results obtained from the calculations previously explained in Chapter Four of this thesis (Section 4.8.2).

5.5.1 Nature of Analytic and Holistic Scores

Data obtained from scores revealed that in both types of assessment, analytic and holistic, the consistency of scores, in other words the Standard Deviation, was lower prior to training sessions than post to training. However, the holistic scores resulted to have a lower standard deviation average than those obtained in the analytical scores post to training (*SD*= 4.44 vs *SD*= .983). This may suggest that teacher participants found it more difficult to cope with the analytical rubric than with the holistic or that the descriptions on the analytic rubric were not clear enough. On the other hand, the significant difference among standard deviations in both types of assessment may also suggest that training encouraged participants to think more thoroughly about the use of each rubric but needed more time to reflect on both types of assessment and their uses.

It is interesting to notice that Sample 4 received the least disperse, as exemplified by the Standard Deviation results, analytic scores prior and post to training (*SD*=3.70, *SD*=4.07) and on the holistic assessment (*SD*=.82, *S*=.84) post to training. This may suggest that this Sample caused the less controversy among teacher participants. This Sample may have

included specific linguistic features that allowed teacher scorers assess this text more consistently. However, the specific features of each text are not in the scope of this study.

Specifically, and in relation to teachers´ analytical scores, scores revealed that Sample 1 received the lowest score mean (*M*=10.60, *M*=10.43) pre and post to training while Sample 5 was given the highest Means of scores (*M*=15.88, *M*=16.52). In terms of the consistency of scores, Sample 4 received the less disperse scores prior and post to training (*SD*= 3.70, SD= 4.07) while Sample 5 received the most disperse scores representing a Standard Deviation of *SD*=4.67 post to training, while Sample 2 was scored most disperse (*SD*= 5.22). In all the samples except two (Sample 1 and 5) SD levels increased meaning that after training scores were more disperse while for Samples 1 and 5 post to training their SD values diminished. These results may support the belief that only in the case of these samples training may have helped teachers therefore considering the nature of the written sample a factor to consider in writing assessment. Table 12 portrays the data here explained.

Table 12 Nature of Analytical Scores

Pre-Training Analytic Scores					**Post-Training Analytic Scores**			
Sample	Max Score	Min Score	M	SD	Max Score	Min Total	M	SD
1	23	2	10.60	4.36	20	2	10.43	4.30
2	25	4	16.15	4.41	23	4	14.87	5.22
3	25	5	18.81	4.23	25	7	18.85	4.28
4	22	7	15.48	3.70	22	5	15.40	4.07
5	24	5	15.88	4.67	25	6	16.52	4.34
Average	23.80	4.60	15.38	4.27	23	4.80	15.21	4.44

In relation to teachers´ holistic assessment, and as shown on Table 13, Sample 1 received the lowest scores (*M*=1.92, *M*=1.94) prior and post to training. On the other hand, Sample 3 received the highest means of scores both prior and post to training (*M*=3.85, *M*= 3.94). In terms of the consistency of scores, Sample 4 received the least disperse scores (*SD*= .82) prior to training and post to training (*SD*=.84). The most disperse scores prior to training were provided to Sample 5 (*SD*=1.08) while post to training Sample 3 received the most disperse scores (*SD*=1.10). It is interesting to notice that in all the cases except two (Sample 3 and 5) increased their SD levels meaning that post to training scores were more disperse. Ideally, scores post to training would be expected to be less disperse meaning that training encouraged evenly spread out scores. However, this was only true for Sample 3 (*SD*=.92 vs *SD*=.88) while Sample 5 maintained value (*SD*=1.08). This result may support my belief that the nature of the written sample may also have an important role in teachers´ assessment: its´ complexity, number of words written, degree of difficulty to assess among others.

Table 13 Nature of Holistic Scores

	Pre-Training Holistic Scores				**Post-Training Holistic Scores**			
Sample	Max Score	Min Score	M	SD	Max Score	Min Score	M	SD
1	4	0	1.92	.89	5	0	1.94	1.01
2	5	1	3.06	.88	5	1	3.12	1.10
3	5	1	3.85	.92	5	2	3.94	.88
4	4	1	3.00	.82	5	1	3.19	.84
5	5	0	2.80	1.08	5	0	3.19	1.08
Aveg	4.6	.60	2.92	.92	5	.80	3.07	.98

5.5.2 Impact of Assessment Training on Analytic and Holistic Assessment

As a hypothesis, it was predicted that training would have a significant impact on the scores teachers provided to sample papers (alternative hypothesis). It was also considered that the training sessions would aid in the improvement of the inter-rater reliability of the scores given to the five sample papers.

A reliability analysis was conducted, specifically a two-way mixed Intra class Correlation Coefficient, to determine the levels of reliability of analytic and holistic assessment prior and post to training of the forty-eight teacher assessors. All scores and measures were analysed considering average measures provided instead of the single measures of scores. When analysing the pre- and post- training analytic scores, results indicated that the Intra class correlation coefficient (ICC) of both scores were *ICC*= 0.957 with *95% confident interval* = 0.880-0.995. As shown on Table 14 and based on the ICC results (ICC>0.7), it can be concluded that the level of reliability was of excellent level (Cicchetti, 1994). This level of reliability was maintained prior and post to training when assessing analytically. ICC calculations on holistic assessment prior and post to training suggested that post training scores were more reliable (*ICC*=.961 with *95% confident interval*= .891-.995) than those provided prior to training (*ICC*=.960 with *95% confident interval*= .889-.995). Thus, on both rounds of assessment teachers´ inter reliability showed to be of excellent level (Cicchetti, 1994).

It can be stated that the alternative hypothesis for holistic assessment was accepted. In other words, training allowed for slightly higher levels of reliability when using a holistic

scoring scale. However, analytic assessment did not show signs of decreasing or increasing reliability. This may suggest that, assessment sessions provided teacher participants with more holistic assessment preparation in comparison to analytic practice. It may also imply that assessment training did not provide enough analytic assessment practice that could allow teachers improve their reliability levels. But, considering that the reliability levels were at the excellent level in all the rounds, it can be concluded that assessment training may not have been a determining factor to cause effect on reliability. Other factors could have been involved in the results obtained from this calculation which still need to be explored.

Table 14 Reliability of Pre and Post Training of Analytic and Holistic Assessment

Intra class Correlation Coefficient (ICC)							
Type of Assessment	Intra class Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower	Higher	Value	df1	df2	Sig
Pre-Training Analytic	.957	.880	.995	36.035	4	184	.000
Post-Training Analytic	.957	.881	.995	47.531	4	188	.000
Pre-Training Holistic	.960	.889	.995	29.201	4	184	.000
Post-Training Holistic	.961	.891	.995	39.401	4	188	.000

When the score means provided prior and post to training were compared by conducting a Paired Sample t-test it was found that, analytically, scores were not significantly different (none of the significance levels represented less than 0.05 p=<0.05) therefore suggesting that training sessions did not impact significantly on the analytical scores. For instance, Sample 3 received the less significant scores (M=-.04, SD=4.82, t (47) =-.06, p=>.05.), therefore rejecting the alternative hypothesis.

When holistic scores provided prior and post to training were compared it also resulted that none of the scores were significantly different. Although all of the Samples received different scores, Sample 1 received the less significant scores (*M*=.02, *SD*=1.24, *t* (47)-.11, *p*>.05) Therefore, suggesting that the difference in holistic assessment done prior and post to training was not of great impact (none of the significance levels represent less than 0.05 *p*=<0.05).

It can be concluded that training sessions did not impact significantly the holistic assessment of the five written samples therefore rejecting the initial alternative hypothesis, which predicted training would impact significantly analytic and holistic assessment of writing. Table 15 depicts the differences found among analytic and holistic scores obtained from the paired t-test.

Table 15 Significance of Pre and Post Training Analytic and Holistic Scores

				95% Confidence Interval of the Difference				
Pairs	M	SD	Std. E M	Lower	Upper	t	Df	Sig 2-tailed
Analytic Scores								
Sample 1	.188	4.26	.61	-1.05	1.42	.30	47	.76
Sample 2	1.27	4.79	.69	-.12	2.66	1.83	47	.07
Sample 3	-.04	4.82	.69	-1.44	1.35	-.06	47	.95
Sample 4	.08	4.03	.58	-1.08	1.25	.14	47	.88
Sample 5	-.64	4.73	.68	-2.01	.73	-.93	47	.35
Holistic Scores								
Sample 1	-.02	1.24	.18	-.38	.34	-.11	47	.90
Sample 2	-.06	1.21	.17	-.41	.28	-.35	47	.72
Sample 3	-.08	1.00	.14	-.37	.20	-.57	47	.56
Sample 4	-.18	1.00	.14	-.47	.10	-1.29	47	.20
Sample 5	-.38	1.40	.20	-.79	.02	-1.90	47	.06

It is my belief that this result could have been encouraged by the small amount of time during which training sessions were conducted. Although two sessions were provided to participants, only 2.5 to 3 hours were spent during each session. I believe that to obtain significant levels of change in scores it is necessary to provide assessors more and constant assessment training. Additionally, since reliability levels were found to be high prior and post to training, other factors could have been present in the reliability obtained. For instance, the proficiency level of scripts may have been too low therefore easier to score. Five samples were given to each participant to score thus the low number of samples that each assessor scored could have also impacted the reliability levels obtained. More samples scored by each participant might have resulted in different values.

5.5.3 Impact of Teachers´ Personal Background on Writing Assessment

With the purpose of analysing the differences in scores of teachers with three different characteristics: teachers´ personal background such as gender, teaching experience and academic background, an Independent Sample T-Test was conducted with the data obtained post to training. The null hypothesis being considered for these personal background characteristics is that gender, teaching experience and academic background have an unequal significance in scores provided to papers while the alternative hypothesis sustains that an equal significance exits.

In relation to gender, the analytic scores of 17 males and 31 females suggested that post to training, males were more lenient in their scores. However, on Sample 4 females were found to be more lenient by providing a higher means of scores (M=15.58 vs M= 15.06). The rest of samples received more lenient scores from male teachers. In terms of the

standard deviation, 4 out of 5 Samples received less disperse scores on behalf of females in the post training scoring round: on Samples 1, 3, 4, and 5 females were found to provide more inconsistent scores (*SD*=4.64, *SD*= 4.66, *SD*= 4.17, *SD*= 4.23).

In regard to holistic assessment, males were more lenient in scores provided to Samples 1, 2 and 4 while females were more severe in the scores provided to these samples. Means provided by males on these samples were higher in relation to the means provided by females. In regard to the standard deviation, males were less dispersed in their scores to Samples 1, 2 and 4 (*SD*=.83, *SD*=.70, *SD*=.94). Table 16 portrays the descriptive data for both males and females previously described.

Table 16 Gender and its Impact on Analytic and Holistic Scores

Post Training Analytic Scores						
Sample	**M**		**SD**		**Std. Error Mean**	
1	M= 10.59	F=10.32	M= 3.72	F=4.64	M= .90	F=.83
2	M= 15.00	F=14.81	M= 5.65	F=5.07	M= 1.32	F=.91
3	M= 20.88	F=17.74	M= 2.54	F=4.66	M= .61	F=.83
4	M= 15.06	F=15.58	M= 3.99	F=4.17	M= .96	F=.74
5	M= 18.46	F=15.45	M= 3.95	F=4.23	M= .96	F=.70
Post Training Holistic Scores						
1	M= 1.76	F=2.03	M= .83	F=1.11	M= .20	F=.19
2	M= 3.00	F=3.19	M= 1.17	F=1.07	M= .28	F=.19
3	M= 4.00	F=3.90	M= .70	F=.97	M= .17	F=.17
4	M= 2.88	F=3.35	M= .92	F=.75	M= .22	F=.13
5	M= 3.54	F=3.00	M= .94	F=1.12	M= .22	F=.20

F= Female M= Male

Once the Independent t-test was conducted in relation to analytic scores, Levine's Test for Equality of Variances proved to show no violations on Samples 1, 2, 4 and 5, *p1*= .349, *p2*=.326, *p3*=.008, *p4*=.720, *p5*=.75. As shown on Table 17, data suggests that the

difference in scores provided by males and females to Sample 3 and 5 were significantly different, *t* (43) =2.56, *p*=<.05 (*p*=0.14), Cohen´s D=.83 and, *t(*43)=2.40, *p*=<.05 (*p*=.020), Cohen´s D=.73 correspondingly. On both Samples, males were considerably more consistent in their assessment, therefore confirming that gender has a significant impact on writing analytic assessment. I consider, that the 95% Confidence Interval Difference is among .49 –5.60. Table 17 exemplifies these values and provides the interval differences. It can be concluded that for Samples 3 and 5 the null hypothesis is rejected thus accepting the alternative hypothesis for Samples 1, 2 and 4.

Very differently from the analytic assessment, among the holistic scores it was found that the means of scores provided by males (ranged from *M*= 1.76, *SD*=.831 to *M*=4.00, *SD*=.70) and females (ranged from *M*= 2.03, *SD*=1.11 to *M*=3.90, *SD*=.97) to all the sample papers were not significantly different. Levine's Test for Equality of variances proved to show no violations on all Samples, *p1*= .42, *p2*=.28, *p3*=.07, *p4*=.94, *p5*=.52. As shown on Table 17, data suggests that the difference in scores provided by males and females to Sample 3 were the least significantly different among all the samples, *t* (43).35, *p*=>.05 (*p*=0.72), Cohen´s D=.12. In this case, the null hypothesis is accepted thus considering that the differences of scores among males and females are not significantly different. Therefore, it can be stated that in the case of holistic assessment gender is not a determining factor to obtain significantly different and consistent scores thus suggesting that holistic assessment may lead to more reliable assessment since gender differences were not of impact.

Table 17 Significance of Gender Impact on Assessment Scores

	Levene's Test for Equality of Variances					95% Confidence Interval of the Difference	
	F	Sig.	T	Df	Sig (2-tailed)	Lower	Upper
Analytic Post Training Scores							
Sam 1	.89	.34	.20	46	.84	-2.37	2.90
Sam 2	.98	.32	.12	46	.90	-3.01	3.40
Sam 3	7.57	.01	2.56	46	.01	.67	5.60
Sam 4	.13	.72	-.42	46	.67	-3.01	1.97
Sam 5	.10	.75	2.40	46	.02	.49	5.52
Holistic Post Training Scores							
Sam 1	.65	.42	-.86	46	.39	-.88	.35
Sam 2	1.15	.28	-.57	46	.56	-.86	.48
Sam 3	3.30	.07	.35	46	.72	-.44	.63
Sam 4	.005	.94	-1.91	46	.06	-.97	.02
Sam 5	.40	.52	1.64	46	.10	-.11	1.17

In regard to teachers´ teaching experience, the 48 participants were divided among two groups with similar characteristics, 1) those that had 5 years of teaching experience or less (33 teachers) and 2) those that had more than 5 years of teaching experience (15 teachers).

Considering teachers' analytic assessment, Sample 5 received the same Mean of scores in both experienced and less experienced teachers (*M*=16.52). However, the more experienced group of teachers were more disperse in their scores, specifically with Sample 2 (*SD*=5.43) while the same group was less dispersed on Sample 3 (*SD*= 4.35). Once again, this difference in standard deviation results could reflect that the characteristics of the Samples could have impacted more the consistency of scores than participants´ training experience. Since more experienced teachers were both more and less disperse in comparison to less experienced ones on different writing Samples, it is my belief that the

length of the text, its level of proficiency or the number of samples assessed could have had a significant impact on scores.

As shown in Table 18, holistic assessment of writing, the most experienced group of teachers, those that have more than five years of teaching experience, provided the most lenient scores on Samples 1, 2 and 5 (*M1*= 1.67, *M2*= 2.93 and *M3*= 3.00 respectively) while Samples 3 and 4 received the most severe scores from the same group teachers (*M3*=4.13, *M4*=3.40). Considering how disperse the scores provided were, the most experienced teachers were more inconsistent in their holistic scores on Samples 1, 2, 3, and 5 (*SD*=1.11, *SD*=1.22, *SD*=.92 and *SD*=1.36) as well as for analytic assessment.

Considering the previously described data, we can consider that novice teachers are more consistent in their analytic and holistic assessment. Maybe in part to their lack of experience which could allow them to see assessment as a simple factor of learning in opposition to more experienced teachers who may tend to involve more factors in their assessment processes thus making it more difficult to attain consistency. This may contradict previously discussed research studies such as those conducted by Lim (2011) and Weigle (1998) where more experienced teachers were more consistent in their scores than their inexperienced peers.

Table 18 Impact of Teaching Experience on Analytic and Holistic Assessment

Post to Training Analytic Scores			
Sam	**M**	**SD**	**Std. Error Mean**
1	a) 10.79 b) 9.60	a) 4.24 b) 4.45	a).74 b) 1.15
2	a) 15.42 b) 13.67	a) 5.12 b) 5.43	a).89 b) 1.40
3	a) 18.55 b) 19.53	a) 4.28 b) 4.35	a).74 b) 1.12
4	a) 15.12 b)16.00	a) 4.18 b) 4.87	a) .72 b) 1.00
5	a) 16.52 b) 16.52	a) 4.13 b) 4.92	a) .72 b) 1.27
Post to Training Holistic Scores			
1	a) 2.06 b) 1.67	a) .96 b) 1.11	a) .17 b) .28
2	a) 3.21 b) 2.93	a) 1.05 b) 1.22	a).18 b).31
3	a) 3.85 b) 4.13	a) .87 b) .91	a).15 b) .23
4	a) 3.09 b) 3.40	a) .84 b) .82	a) .14 b) .21
5	a) 3.27 b) 3.00	a) .94 b) 1.36	a) .16 b) .35

A= Five years or less **B= More than Five years**

In relation to the significance of the difference among analytic scores within experienced and inexperienced teachers, the Independent T-Test allowed the researcher identify that these differences were not significant. Once the Independent t-test was conducted, Levine's Test for Equality of variances proved to show no violations on all the Samples, *p1*=.68, *p2*=.96, *p3*=.63, *p4*=.84, *p5*=.26.

As shown on Table 19, data suggests that the difference in scores provided by novice (Group A) and experienced teachers (Group B) to Sample 5 demonstrated that the differences among Group A and B were not significant, t (46) =.00, p=>.05 (p=.99), Cohen´s D=0. I consider, that the Confidence Interval Difference is within 95% correct thus meaning that the difference found among scores is not significant enough to conclude that more experienced or less experienced teachers are more accurate in assessing analytically their students' work.

In regard to the differences among both groups of teachers and the significance of the holistic scores provided, it was found that all the scored samples received the value p=>.05 (as shown on Table 19, specifically 2-tailed significance). Once again, Levine's Test for Equality of Variances proved to show no violations on all the Samples, *p1*=.95, *p2*=.54, *p3*=.96, *p4*=.97, *p5*=.16. As shown on Table 18, once again (as in analytic assessment) Sample 5 along with Sample 2 received scores that demonstrated that the differences among Group A and B were not significant, *t*(46)=.80, *p*=>.05 (*p*=.42), Cohen´s D= .23 thus indicating that the impact of teachers´ teaching experience on analytic and holistic assessment is not significant enough to conclude that it may have a role in the assessment of writing. In other words, the null hypothesis, which considered that there was not a relationship among teaching experience and scores obtained, was accepted. Therefore, it can be argued that teachers´ teaching experience is not relevant to the reliability of writing assessment but instead other factors such as training or the nature of the written sample may have a more determining role in assessment.

Table 19 Significance of Teaching Experience Impact on Analytic and Holistic Assessment

	Levene's Test for Equality of Variances					95% Confidence Interval of the Difference	
	F	Sig.	T	Df	Sig (2-tailed)	Lower	Upper
Analytic Post Training Scores							
Sam 1	.16	.68	.88	46	.38	-1.51	3.89
Sam 2	.002	.96	1.08	46	.28	-1.51	5.03
Sam 3	.23	.63	-.73	46	.46	-3.68	1.71
Sam 4	.04	.84	-.68	46	.49	-3.44	1.68
Sam 5	1.30	.26	-.00	46	.99	-2.75	2.74
Holistic Post Training Scores							
Sam 1	.003	.95	1.24	46	.21	-.24	1.02
Sam 2	.36	.54	.80	46	.42	-.41	.97
Sam 3	.002	.96	-1.03	46	.30	-.83	.26
Sam 4	.001	.97	-1.18	46	.24	-.83	.21
Sam 5	1.97	.16	.80	46	.42	-.41	.95

The third teacher characteristic considered in the Independent T-test was the academic preparation that participants had. Two main groups were found among the forty-eight participants, 1) those that had a bachelor´s degree and 2) those that were BA students at the moment of the study. It was considered as a null hypothesis that academic background does not have a significant impact on their analytic and holistic scores. On the other hand, the alternate hypothesis would indicate that teachers´ academic background influences significant differences provided by these groups.

As depicted graphically in Table 20, analytic scores provided by those with an undergraduate degree (Group B) indicated that these are more lenient with their scores since all the Sample papers of this group received the highest means (ranging from *M*=20.37 to *M*=10.57). However, teachers who already had a degree provided less disperse scores on Samples 2, 3, 4 and 5 (*SD2*=4.77, *SD3*=3.56, *SD4*=3.74 and *SD5*=3.95). Therefore, suggesting that those that have completed a degree may produce more levels of consistency among analytic assessment.

In relation to holistic scores and similarly to analytic assessment, Group B provided the most lenient scores to all the Sample papers (*M1*= 197, *M2*=3.17, *M3*=4.23, *M4*=3.40, *M5*=3.27). On the other hand, the most disperse sets of scores was provided by an undergraduate student on Sample 5 (*SD*= 1.14) while the least disperse scores were provided to Sample 3 by a member of Group A (*SD*= .784). Among holistic assessment these results may suggest that BA students have more difficulty coping with the analytic rubric therefore resulting in more consistent assessment with the use of a holistic scoring tool on three (Sam1, Sam 3, Sam 5) of the five Sample papers.

Table 20 Teacher Academic Background and its Impact on Assessment

Post to Training Analytic Scores						
Sam	**M**		**SD**		**Std. Error Mean**	
1	a) 10.17	b) 10.57	a) 4.24	b) 4.40	a) 1.00	b) .80
2	a) 14.39	b) 15.17	a) 6.02	b) 4.77	a)1.41	b).87
3	a) 16.33	b) 20.37	a) 4.28	b) 3.56	a)1.01	b) .651
4	a) 14.17	b) 16.13	a) 4.39	b) 3.74	a) 1.03	b) .68
5	a)15.56	b)17.09	a) 4.90	b) 3.95	a) 1.15	b) .721
Post to Training Holistic Scores						
1	a) 1.89	b) 1.97	a) 1.02	b) 1.03	a) .241	b) .189
2	a) 3.06	b) 3.17	a) 1.25	b) 1.02	a).297	b).186
3	a) 3.44	b) 4.23	a) .784	b) .817	a).185	b) .149
4	a) 2.83	b) 3.40	a) .857	b) .770	a) .202	b) .141
5	a)3.06	b)3.27	a) .998	b)1.14	a) .235	b) .209
	a) BA Student	**b) Undergraduate Degree**				

Considering these characteristics an Independent Sample T-test was run to determine if the differences among the scores of these two groups of participants were significant or not. As shown on Table 21, Levene´s Test for Equality of Variance showed no violations (values ranging from p=.26 to p=.96, all above p=.05). Test results suggested that teachers´ academic background impacted significantly on scores provided analytically to Sample 3, $t(46)$= -3.51, p=<.05 (p=.001), Cohen´s D=1.02. In other words, participants who had an undergraduate degree *(*M=20.37, SD=3.56) had more consistent and significant analytic scores thus suggesting that participants who were more academically prepared had more impact than those who were studying their BA program (M=16.33, SD=4.28). These findings may suggest that teacher professional development, such as that of obtaining a university degree, may lead to more accurate and reliable writing assessment. However, factors such as teachers´ age, teachers´ assessment preference or their teaching and assessment experience need to be considered as well.

Table 21 Significance of Teachers´ Academic Background on Analytic and Holistic Assessment

	Levene's Test for Equality of Variances					95% Confidence Interval of the Difference	
	F	Sig	T	Df	Sig (2-tailed)	Lower	Upper
Analytic Post Training Scores							
Sam 1	.05	.81	-.30	46	.75	-3.00	2.20
Sam 2	1.10	.29	-.49	46	.62	-3.94	2.38
Sam 3	.15	.69	-3.51	46	.001	-6.34	-1.72
Sam 4	.17	.67	-1.64	46	.10	-4.36	.43
Sam 5	3.36	.07	-1.19	46	.23	-4.13	1.05
Holistic Post Training Scores							
Sam 1	.08	.77	-.25	46	.80	-.69	.54
Sam 2	1.37	.24	-.33	46	.74	-.78	.55
Sam 3	.00	.98	-3.28	46	.00	-1.27	-.30
Sam 4	.01	.91	-2.36	46	.02	-1.04	-.08
Sam 5	.28	.59	-.64	46	.52	-.86	.44

On the other hand, when conducting the analysis of the holistic scores it was found that Samples 3 and 4 were provided with scores that were significantly different. On Sample 3 (*M*=3.44, *SD*=.784) and 4 (*M*=2.83, *SD*=.85) BA students were those who scored more consistently over those that had an academic degree.

As Table 21 portrays, Levene´s Test for Equality of Variance showed no violations (values ranging from *p*=.24 to *p*=.98, all above *p*=.05). Test results suggested that teachers´ academic background impacted significantly on scores provided holistically to Sample 3, *t* (46) =.98, *p*=<.05 (*p*=.02), Cohen´s D=.98 and Sample 4 *t* (46)= -2.36, *p*=<.05 (*p*=.02), Cohen´s D=.69. In other words, participants who were BA students *(Sample 3 M*=3.44, *SD*=.78 and Sample 4 *M*=2.83, *SD*=.85) had more consistent holistic scores thus suggesting that BA students had a more significant impact on holistic scores.

5.5.4 Section Conclusion

This section attempted to answer RQ5 *(To what extent does writing assessment training and teachers´ personal and academic background impact their use of analytic and holistic scoring tools to assess written texts in the EFL classroom?*), which focuses on the quantitative analysis of the analytic and holistic scores that forty-eight Mexican EFL university teachers gave to five opinion essay samples. It also sought to understand the role that gender, teaching experience and academic background may have on participants´ analytic and holistic assessment. A reliability analysis, specifically an Intra Class Correlation Coefficient (ICC), a Paired Sample t-test, and an Independent Sample t-test were run on the SPSS v.23 software program with the scores that the forty-eight teachers provided to the Samples.

The results obtained from the ICC calculations suggested that inter assessor reliability on both analytic and holistic assessment was at the excellent level (Ciccetti,1994) both pre- and post-training. Additionally, it was found that holistic assessment increased its reliability levels post-training among the forty-eight participants. These findings may suggest that other factors in addition to assessment training, such as the nature of the writing sample, the number of samples scored, the amount of time taken to score the samples, teachers´ teaching and assessment experience among other factors need to be analysed to explore possible aspects that have an impact on writing assessment reliability.

Calculations from a Paired Sample t-Test were run to compare the analytic and holistic scores obtained pre and post to training from the forty-eight teachers. Results suggested

that training sessions did not have a significant effect on the analytic and holistic scores provided to the five written Samples. This could mainly be attributed to the little amount of time dedicated to the training provided. Two assessment-training sessions with an approximate duration of 2.5-3 hours were provided. It is possible that teachers needed more time to reflect on the contents and practice provided during the sessions but above all to change their assessment process. These results could also yield the need to have more samples scored by participants with the intention of obtaining a clearer picture of the effects of training as well as more participants that could provide a bigger sample of scores to analyse.

An Independent Sample t-test was conducted considering three teacher characteristics: gender, teaching experience and academic background. Data suggested that gender had a significant role in analytic assessment. More specifically, males resulted to have more consistent scoring while using an analytic scoring scale therefore suggesting that holistic assessment is a more reliable and a fairer type of assessment since gender was not a factor of impact. Additionally, academic background was also found to have a significant impact on scores, specifically on analytic assessment. Results pointed out that teachers who had an undergraduate degree were more consistent in their analytical assessment while undergraduate students were more consistent with holistic assessment. These results may also lead to argument that analytic assessment is a more subjective and difficult type of assessment since it is the type that is impacted the most by external factors such as gender and academic background of assessors.

Finally, analysis of scores suggested that the teaching experience of an assessor is not a factor of significant effect on assessment reliability. In other words, either experienced or novice, the difference among pre- and post- training analytic and holistic scores that teachers provided to the five samples were not of significant impact. These results may suggest that although teacher experience did not impact assessment, language assessment experience and previous language assessment literacy experience may need to be further considered as factors that may have a significant difference in the reliability of scores. It may also be argued that more than a need for teaching experience, teachers need experience assessing language skills to ensure the reliability and validity of assessment. The following section provides a discussion of these and other important findings.

Chapter 6: Discussion

The present study had the main purpose of analysing the impact that assessment training had on EFL teachers´ reported classroom assessment of students´ writing skills in three main areas of impact, 1) reported writing assessment practices in the classroom, 2) teachers´, program managers´, students´ perceptions of writing assessment and of writing assessment training and 3) teachers´ use of scoring tools to provide a holistic and analytic score to students´ texts.

One of the main findings of this study was that writing assessment training (WAT) has greater impact on teachers´ meta-analysis skills in comparison to their classroom assessment of writing. In other words, the analysis of themselves and their assessment activities as teachers and as assessors in their classroom. This finding may suggest that experiencing training sessions with a group of peers that face the same contextual difficulties sets forward the importance of socialization in assessment literacy (Scarino, 2013, 2017; Lam, 2015; Koh *et al.,* 2017). This could mean that when teachers have opportunities of sharing with others their difficulties when assessing, it encourages the understanding of their own knowledge of assessment and makes way for the understanding of new knowledge (Scarino, 2013, 2017) thus triggering their self-awareness skills.

Results also suggested that WAT triggered the teaching of writing, writing assessment awareness and its importance to students´ language development (Crusan, 2010; Weigle, 2007). In other words, it gave teachers the opportunity to reflect on the importance that the

teaching of writing has for language students and how it´s assessment can encourage students to give importance to it. Additionally, WAT encouraged change in teachers´ and language managers´ perceptions of writing instruction and assessment which lead one manager and one language teacher to propose innovations to the assessment procedures in their institutions, improvement in the assessment process and the update of scoring tools used (Huot, 2002; Scarino, 2013), such as was the case of PM4 and TP16.

In regard to the quantitative calculations, results suggested that when comparing pre- and post to WAT reliability levels of both holistic and analytic assessment, they were maintained at excellent level, being holistic assessment the most reliable. Data depicted that teachers´ academic background (undergraduate degree or undergraduate student) had an effect on the scores provided: those that finished a degree were more consistent in their analytical assessment while those that were studying at the moment of the study provided more reliable holistic assessment. It was also found that gender and teaching experience caused an impact on scores but these differences were not considered significant as explained in Table 17 and Table 19 of this document. The following sections focus on the description of the specific changes that participants of this study reported to have experienced and the possible implications that these results may suggest.

6.1 The Writing Assessment Training Impact Categorization

Qualitative data that emerged from data analysis, as described in Chapter Five of this thesis, allowed the proposal of the Writing Assessment Training Impact Categorization

(WATIC, Figure 7) which is included below. This Figure represents the reported impact of WAT on teachers´ teaching and assessment practices in this specific Mexican EFL context, in an attempt to acknowledge the importance of contextual factors (Crusan, 2010; Huot, 2002; White, 1990; Fulcher and Davidson, 2007, Yan, Fan and Zhang, 2017) such as institutional policies or the nature of the EFL program.

Figure 7 Writing Assessment Training Impact Categorization (WATIC)

The WATIC is a three-level assessment impact construct which portrays the type of impact that training caused in teachers in three broad areas. Each level was constructed from the themes, subthemes and categories that emerged in the qualitative analysis of data, which followed a grounded theory approach. The first level includes the three major areas of impact such as Writing in the EFL Classroom, Classroom Assessment of EFL Writing and Teacher Self-awareness (Level 1). Each area is divided in two to five different subthemes (Level 2) which represent the different types of impact found within each area. Area 1 `Writing in the EFL Classroom´, was divided in two subthemes: A) Writing Activities and B) Feedback Techniques. The subthemes Assessment Procedures and Scoring Tools were included in Area 2 titled `Classroom Assessment´ while Area 3 `Teacher Self-Awareness´ depicted five different subthemes: Nature of Writing, Teaching of Writing, Assessment Procedures, Writer Stance and Student Stance. Each subtheme represents the actions that participants of the study reported they had conducted in their classroom and that represented, in the TPs´ perception and my own (as the main researcher) interpretation, the effect of training in their practice or their views in regard to writing assessment. Each subtheme portrays from two to five different categories which all were compiled in Level 3 of the WATIC, as can be identified in Figure 7.

It is worth pointing out that the area that resulted the largest in regard to the number of subthemes and categories was Area 3 ` Teacher Self-awareness´. It can be argued that assessment training had the most impact on teachers´ reflection of their own perceptions and assessment procedures even though training was delivered in short amounts of time on a limited number of days.

The WATIC may serve as a guide for teachers´ to reflect on their own strengths of their assessment processes and make decisions as to what needs to be done to improve their weaknesses. It may also be useful for FL program managers and language institute administrators to visualize the potential benefits of providing their staff with WAT thus to make decisions regarding the specific training that is cost and time feasible for them.

6.2 Reported Classroom Assessment Practices

Data analysis revealed that four of the eleven teacher participants (TPs) of phases four and five experienced distinct types of changes in reference to their regular assessment procedures while others reported no specific impact in their classrooms. These participants instead reported a change in their self-awareness (third main theme identified and portrayed in Level 1 of the WATIC), of the nature of writing the importance of teaching writing, the importance of writing assessment and their stance as a writer and as an EFL teacher (Scarino, 2013; Lam, 2015; Koh *et al.*, 2017).

Those that reported to have experienced an actual change in their classroom assessment (TP22, TP62 and TP313) or their use of scoring tools (TP22, TP62, TP313 and TP315) explained that minor changes conducted included a redefinition of assessment purposes, an inclusion of students in the assessment process (Leung and Mohan, 2004) and an improvement of the assessment process followed, as portrayed in the second main theme of the WATIC: Writing Assessment in the EFL Classroom. On the other hand, all the teachers except one (TP326) reported that assessment training had encouraged them to analyse and be more self-aware of how to improve their own teaching and assessment of EFL writing (TPs with the least professional background) while it allowed others to update

assessment practices they already knew (TPs with more ELT academic background). These results are supported by the Paired Sample t-test calculations conducted in this study in which it was found that WAT did not significantly impact the analytic and holistic scores teachers gave to the five opinion essay samples. Therefore, confirming that impact of WAT is minor in teachers´ classroom assessment (Koh *et al.,* 2017), but beneficial for other aspects of teachers´ assessment literacy such as their conceptualizations and interpretations of assessment (Scarino, 2013, 2016). This finding may suggest that training had a positive effect on participants´ assessment behaviour in their classroom and their beliefs towards the importance of assessment rather than on the improvement of the quality of assessment. Positive impact on classroom assessment as an effect of training or workshops may be actually hard to obtain and is rarely measured (Jin and Jie, 2017), however it is necessary to understand if these assessment literacy tools can actually improve the quality of teachers´ language classroom assessment.

The results of this project may relate to those found by Elder *et al*., (2005) who focused on analysing the perceptions that eight experienced raters of English diagnostic writing had of the feedback provided to their scoring processes as part of their online training. The researchers found that feedback was perceived as useful for raters´ practice. Even though the study was carried out in a different research context (ESL large-scale testing), Elder *et al*., (2005) described that feedback provided also encouraged participant awareness of their own assessment behaviour. In this study, feedback was not provided to participants but instead group discussions during training allowed teachers to reflect on their own classroom assessment processes and initiate planning on how to improve their assessment

processes. According to TP32 and TP37, assessment training should be constant and permanent thus suggesting that training may be of more benefit to their practice if these characteristics are complied with. Researchers (Roux and Valladares, 2014; Koh *et al.,* 2017) have suggested the implementation of follow-up measures, such as permanent training or reflective sessions, supported by educational institution authorities that may provide teachers with the opportunities to improve their practice.

It was surprising to find that training was successful in raising awareness of the importance of writing as a language skill and that through its classroom assessment, stakeholders can learn to also give writing its place in the EFL classroom and the EFL curriculum. Additionally, the proposed Categorization, the WATIC (Figure 7, found on pg. 237 of this document), may allow teacher trainers and school decision makers´ to picture some of the benefits of providing training to their staff. It is also an example of the difficulty of providing teachers with training. Specifically, the issue of being uncertain of the exact benefit that training may result in teachers´ actual classroom assessment and the large amounts of time needed to actually identify improvement as a direct result of assessment training.

The WATIC may contribute to the Assessment Literacy Inventory (ALI) developed and validated by Metler and Campell (2005). While the ALI focuses on the assessment standards of the American education system and its correlation with classroom teachers´ assessment literacy, and their perceptions of their assessment literacy (Metler, 2003; Metler and Campell, 2005) the WATIC proposed in this study focuses on Mexican EFL

teachers´ who reported changes in their classroom assessment procedures therefore an initial step to contribute to the development of the complex area of assessment literacy.

6.3 Impact of Writing Assessment Training on Language Programs

Data obtained from Institutions A, B, C and D, revealed that participants who work in the same institution and under the same conditions continued to assess writing very differently and some did not consider it worth assessing (Institution A) therefore suggesting that an institutional culture (Chen *et al.,* 2013) is not followed nor is the socialization aspect of assessment (White, 1990; Scarino, 2013; Chen *et al.,* 2013) considered. In other words, this finding suggested that teachers did not share with their peers or their program managers their assessment beliefs, difficulties, or procedures; or that teachers did not homogenously follow their institution´s culture since they assessed students´ writing abilities on their own, following their own procedures, their own assessment purposes and/or their own assessment criteria. On the other hand, in other institutions (Institution B and C) teachers have similar assessment procedures: used the same test, the same assessment activities and the same analytic and holistic scoring criteria. However, they differed in specific actions such as the level of student involvement in the process, the amount of activities used to assess writing and the interpretation given to the tools used to assess the texts (Inbar-Lourie and Donitsa- Schimdt, 2009). These differences of the impact of assessment training within the same institutions and among different institutions suggest that although the assessment of writing is very much influenced by the social context (White, 1990, Chen *et al.*, 2013) in which it is embedded (assessment regulations of the institution, assessment procedures of other teachers, scoring tools used, students´, teachers´ and managers´ beliefs of assessment, perceptions of the potential washback of

assessment among other factors) the individual differences that the teacher may carry could make a big difference in the assessment outcome.

Another factor that could have led to the different approaches and interpretations (Hamp-Lyons, 1990; Grabe and Kaplan, 1996; Weigle, 2002; Weir, 1990; Bachman and Palmer, 2010) that each teacher gave to the scoring scales and assessment processes within an institution is the amount of time teachers had been working with specific assessment tools (Institution D). Although it was reported that assessment training did not provide the institution with an immediate change in their assessment procedures and tools, it did allow teachers and language coordinators to reflect on the need to update the tools used and assessment processes followed.

The remaining three institutions did not implement homogenous assessment procedures therefore teachers chose which type of assessment and assessment tool best fit their practice. This amount of freedom may allow teachers become critical about their assessment of writing and more open to change in relation to those who are imposed specific assessment standards. However, too much liberty without having the academic and theoretical support for choosing a specific assessment method may lead to unreliable and invalid assessment.

6.4 Teachers´ Perceptions of Classroom Writing Assessment and Writing Assessment Training

Regarding teachers' views, data revealed that their changed perceptions referred to four main areas: 1) perceptions of writing assessment procedures, 2) perceptions of writing

assessment scoring tools, 3) perceptions of writing assessment training and 4) perceptions of themselves as EFL classroom assessors.

Regarding the first category, it was found that most of the participants gave more importance to the assessment of the skill and the use of scoring tools after experiencing training. Additionally, given the context in which EFL is embedded in the north-eastern region of Mexico and the experience that teachers reported to have with analytic scoring tools, more teachers reported to prefer using analytic scoring tools in their classrooms while others preferred using a combination of analytic and holistic scoring depending on their assessment purposes and students´ needs (Cumming, 2001; Cheng *et al.,* 2004).

In relation to participants' perceptions of assessment training, the majority believed that training sessions had been practical, useful (Elder *et al.,* 2005, Knoch, 2011), beneficial and supportive for their practice. However, they also believed that two sessions were not enough to actually change and improve assessment practices in the classroom (Koh *et al.,* 2017). These results were supported by those found by the Paired Sample t-test calculations which revealed that analytic and holistic assessment did not receive significant impact post to training.

Data in the present study also suggested that perspectives in terms of the needs of teachers and the content of training should include more assessment practice accompanied by practice in the creation and adaptation of scoring tools (Hasselgreen *et al.,* 2004; Nier *et al.,* 2013; Vogt and Tsagari, 2014). Teachers´ and program managers´ perceptions in

regard to WAT agreed on the need to include more sessions to allow teacher reflection, more time to include thorough assessment practice, and the inclusion of scoring tool construction and adaptation (Esfandari and Myford, 2013). An additional factor added by the managers, was the need to sequence WAT to first approach the teaching of writing as a skill then move on to writing assessment, in other words the need to establish an explicit link between the learning of the skill and the assessment of writing. These results add to those found by Hasselgreen *et al.,* (2004), Nier *et al.,* (2013) and Vogt and Tsagari (2014).

In Hasselgreen *et al.,* (2004), 197 teachers and teacher trainers responded to an online questionnaire and expressed they considered training should focus on creating assessment tools, use of portfolios, peer/self-assessment, interpreting results, establishing validity and reliability throughout statistics, rating student performance in productive skills among others. Nier *et al.,* (2013) found that perceptions in relation to online training included the need to add the practice of assessment tasks and the inclusion of context specific assessment procedures. Therefore, the results of this study correspond to those found by Hasselgreen *et al.,* (2004) and Nier *et al.,* (2013) since teachers considered they needed more scoring samples to practice with and benchmark papers to reflect on (Volante and Fazio, 2007; Lam, 2015).

Finally, results depicted that some teachers had negative while others had positive perceptions (Lopez Mendoza and Bernal Arandia, 2009) of writing assessment. Fear and discomfort were the negative views that teacher felt (Coombe *et al.,* 2012; Stiggins, 1995) after experiencing training while positive views involved an increase of self-confidence, and motivation towards assessment. These perceptions enlighten the path to understand

how teachers perceive themselves, in this case as writers and assessors and how this projection can help teacher trainers and EFL language program managers improve training sessions to approach the specific needs of EFL teachers. It is my personal belief that perceptions encourage and guide improvement of performance, in the ELT area and any other area of study. On the other hand, it is important to emphasize that WAT triggered self-awareness skills in teachers to an extent to which participants were criticizing themselves and their teaching/assessment skills (Scarino, 2013, 2016). Hopefully, these reported self-assessment activities of teachers´ classroom practice and the importance they give to the skill is the initial stage of the improvement of their own assessment of writing in their EFL classroom.

6.5 Language Program Managers´ Perceptions of Classroom Writing Assessment and Assessment Training.

The second group of participants of this study were four language program managers who, at the moment of the study, were heads of the language departments of the institutions under analysis. Results indicated that while two managers (PM1 and PM3) reported to have experienced an impact in their perception of assessment and the actual assessment procedures of the program they managed, the remaining two indicated that more time was needed (Hasselgreen *et al.,* 2014; Nier *et al.,* 2013; Esfandari and Myford, 2013; Vogt and Tsagari, 2014) to actually cause impact and change in assessment.

PM3 explained that she had been able to analyse and self-assess how teachers were being required to assess language skills in their institute (institutional language assessment

policies) therefore considering a minor change to the writing assessment tasks included in the monthly tests at all the proficiency levels. These results may indicate that although the actual effect of WAT on the assessment procedures of an institution may be shallow, the initial stage of change or innovation may be stakeholders´ analysis and self-reflection.

On the other hand, it was interesting to find that teachers, students´ and PMs´ perspectives converged considering that updating experienced teachers´ skills was one of the main benefits of providing training to teachers. PM2 and PM4 reported to have had a positive experience during the sessions without an evident change in the assessment procedures of the program or of their own classrooms. However, they were eager to point out that the main difficulties that were faced as an administrator with the inclusion of writing in the EFL program were the lack of time (Crusan, 2014) and the lack of writing abilities that students presented in their L1. These findings may lead to the argument that WAT may also serve as a tool to find agreement among participants in the benefit of the EFL curriculum and the assessment procedures of the institution.

Consensus among teachers, students and managers was found in terms of the time required to teach and assess writing. However, disagreement in terms of the students´ poor writing skills hindering the assessment of writing (Malone, 2013) was also identified. In my experience, for some non-writing students, learning to write in a foreign language may be difficult. But for others, it may be easier to write than in their first language. Therefore, I consider that this depends on many factors that may actually be in the hands of the teacher and the language manager to address (meeting students´ interests, students´ social context,

and/or students´ extra-curricular activities). Additionally, I believe that managers are the core of suiting language programs to the needs and interests of students. Managers are also co-responsible (shared with teachers) for providing instructors with opportunities of being professionalized (Bailey and Brown, 1996; Metler, 2003; Metler and Cambell, 2005). However, teachers need to also be interested in updating and improving their own practice.

It can be concluded that the actual impact that assessment training may or may not have on teachers´ classroom assessment and on EFL programs depend not only on the instructor of the training sessions but also on teachers´ interest to improve (Roux and Valladares, 2014), teachers´ availability, teachers´ teaching style, institutional support to provide training to staff, and finally institutional culture towards professional improvement. This conclusion may add to Chen *et al.* (2013) claims in which it is stated that EFL teachers´ classroom practices and program managers´ decision making strongly depend on the assessment culture of the institution in the Chinese context. The findings of the study described in this thesis suggest that WAT impact on classroom assessment procedures also depend on the sociocultural conditions in which they operate. Quantitative analysis of data conducted in this study support this finding in the sense that it was found that teachers´ personal characteristics (modelled and shaped by sociocultural factors present in the teachers´ assessment context) such as academic background have an impact on their assessment of writing.

While it can be argued that assessment training did not impact significantly participants´ regular classroom assessment procedures, findings depicted WAT as a powerful tool that

may encourage teachers to be aware and reflect on their practice which may consequently trigger positive impact on writing assessment (Scarino, 2013; Moss, 1996, 2004). Additionally, it may also be a persuasive tool to raise awareness of the importance of writing instruction and assessment in the EFL classroom.

6.6 Student Perceptions of Classroom Writing Assessment

Students, on the other hand, also agreed with their teachers´ perception of the importance of writing to their language development. Students were always aware of this importance (pre- and post to training). Teachers and students reported they felt less afraid of assessing writing (in the teachers´ case) or being assessed (in the students´ case) now that teachers had experienced WAT. Thus, it can be concluded that WAT allowed for teachers, and consequently students, to feel more comfortable and familiar with the development of writing in the classroom and its assessment causing a domino effect on students´ perceptions of writing.

Studies such as those conducted by Crossman (2007) in L1 contexts acknowledge the importance of considering students´ perceptions of assessment. The researcher points out the need to provide students with opportunities to express their beliefs, feelings and attitudes towards assessment during the assessment process. Thus, the results of this study agree with those found by Crossman (2007) in a sense that by acknowledging students´ feelings toward assessment, language development and language learning are more meaningful.

Student participants also had their own perspective of the importance of teacher assessment literacy. All of the students of the participating institutions considered training was essential for every teacher to connect their classroom assessment with reality (students of Institution C) and to update skills of those experienced teachers (students of Institution D). Hence, agreeing with PM2 and PM3 that WAT is useful to update skills of those experienced teachers and to provide new practice to those inexperienced teachers. Students additionally expressed distinct feelings towards the assessment of writing and the criteria used by their teacher to assess writing (Donald and Denison, 2001). While participants of Institution B experienced negative perceptions towards their teachers´ assessment processes (perceived language development was hindered, felt uneasy, uncomfortable) (Sambell *et al.,* 1997), those studying at Institutions C and D expressed positive feelings (comfortable, supported, secure). Specifically, students from Institution D felt that assessment made them feel they were losing their time in pointless assessment criteria that did not allow them to improve their language (Sambell *et al.,*1997). Thus, bringing forward the importance of student perceptions of assessment processes to the development of their language skills (Sambell *et al.,*1997). These perceptions were reported to be consistent throughout the study (pre- and post training). Contrary to the case of Institution A in which students perceived negatively their teachers´ assessment procedures before training, then positively post to training.

It can be argued that at least in the case of the participants of Institution A, assessment training had a positive impact on their teachers´ assessment procedures and use of tools therefore encouraging positive change in students´ perceptions. It is worth commenting

that the fact that students of Institution C requested their writing score be accompanied by specific feedback that could guide their improvement (Randall and Zundel, 2012), may suggest that students are eager to interact with the teacher to improve their language skills. I consider that more in depth analysis of students´ perceptions of assessment is needed.

6.7 Impact of teachers´ personal/academic background and assessment training on analytic and holistic assessment

Individual differences such as age, years of teaching experience and academic background may have an important role in professors´ teaching and assessment practices in the EFL classroom (Weigle, 1998; Eckes, 2008; Contreras *et al.,* 2009; Lim, 2011; Barkaoui, 2011; Esfandari and Myford, 2013, Attali, 2015) thus the importance of its analysis in this project. Results from Independent Sample t-test calculations suggest that most experienced teachers provided more disperse scores both in analytical and holistic assessment, however the differences among scores were not significant enough to conclude that this variable has an impact on writing assessment.

It was surprising to find that the more experienced group of teachers (five years and more of teaching experience) provided more inconsistent analytic and holistic assessment than their novice peers (less than five years of teaching experience). When comparing holistic and analytic assessment less experienced assessors were found to be more consistent (Lim, 2011) in their holistic assessment (Barkaoui, 2007). These conclusions may contradict those pointed out by Weigle (1998) and Attali (2015). Weigle (1998), after analysing sixteen raters (eight experienced and eight inexperienced) and their scores to texts post to training concluded that inexperienced raters are more severe and less consistent in their

ratings than the experienced raters before training. It is stated that although training does not guarantee consistency it can encourage raters to be more self- consistent.

Finally, Attali (2015) after comparing the scores provided to more than 20 written papers pointed out that experienced and non-experienced raters did not differ significantly in the reliability and validity coefficient. However, the results of this study agree with both Weigle and Lim that training triggers auto analysis of teachers´ own assessment procedures and provides opportunities for them to construct their own interpretations of their assessment practices (Weigle, 1998; Lim, 2011; Scarino, 2013, 2016).

Teacher academic background was found to be a significant variable in the analytic and holistic assessment process. Independent t-test calculations conducted suggested that participants who held an undergraduate degree diploma provided less disperse analytical scores while those that were BA students at the time of the study assessed holistically less disperse than their graduated peers. These differences were found to be significant therefore allowing to conclude that teacher academic background has a significant impact on holistic and analytic assessment of writing.

Data obtained in this study strongly suggests that the scores provided by the forty-eight teachers to the five written samples prior and post to training were not significantly different. In other words, the differences of scores does not suggest training impacted their actual scoring of papers. Reliability tests (Intra class correlation coefficient) suggested that inter reliability among teachers´ holistic assessment improved post to training (*ICC*=.961 with *95% confident interval*= .891-.995) than those provided prior to training (*ICC*=.960

with *95% confident interval*= .889-.995). Analytic assessment was found to be at excellent levels (Saxton *et al.,* 2012; Contreras *et al.,* 2009), according to Cicchetti´s (1994) reliability levels, prior and post to training thus it can be considered that training did not impact the reliability of this assessment.

Ideally, teacher trainers want their trainees to improve practice after delivering their sessions. In this study, even though the reliability levels were already at an excellent level prior to training, change post to training was still intended. This was only possible in the case of holistic assessment. This may be attributed to several factors, a) content of training may need to focus more on analytic assessment and over longer periods of time; b) two sessions of WAT may not be enough to get acquainted with analytic assessment (Hasselgreen *et al.,* 2004; Esfandari and Myford, 2013); c) some of the teacher participants pointed out to prefer holistic assessment therefore influencing their more reliable holistic assessment post to WAT. This may also lead to the argument that the nature of the scoring scale and the context in which it is used has a strong influence on its reliability levels of scores. Finally, the small number of samples scored by the participants (five opinion essay samples) may have had an impact on the little difference of reliability levels of pre-and post-scores. By scoring more samples, a wider perspective of this impact may have been obtained.

The results yielded by this study regarding reliability levels among analytic and holistic assessment agree with those found by Barkaoui (2007) in which a statistical comparison among analytic and holistic rubrics was developed. The researcher found that, in University large-scale writing assessment, holistic scoring was more reliable than analytic

scoring due to the number of categories included in the analytic rubric. In this study, the same finding is signalled in the classroom assessment context of writing. Gender was found to impact significantly analytic assessment but not holistic therefore suggesting that holistic scoring may be more reliable since it is not impacted by raters´ gender. Nonetheless, the specific reasons for this outcome were out of the scope of this study.

The data obtained from the paired sample t-test calculations (comparison of analytic and holistic means prior and post to training) is considered to support the findings of the qualitative data in terms of the minimum impact that assessment training was found to have on teachers´ regular writing assessment. Nevertheless, evidence seems to suggest that the major gains of assessment training in teachers was a higher degree of awareness of the importance of writing in the EFL classroom, increase in students´ awareness of the importance of the inclusion of writing in classroom assessment, and an increase in teachers´ reflection in regard to their performance as a teacher and as an EFL assessor (Fulcher, 2012; Scarino, 2013; Koh *et al.,* 2017), traits graphically represented in Figure 7 of this thesis (Chapter 6). This may lead to the claim that providing training to teachers may be the initial stage of teacher development. By providing training on several occasions, teacher participants of this study were encouraged in a deeper reflective process that hopefully will lead to the future improvement of their writing assessment.

Finally, it is crucial to point out that another finding of this study was that teacher academic background significantly impacts analytic and holistic assessment. Independent sample t-test calculations revealed that teachers who held an undergraduate degree diploma had significant impact in their analytic assessment while those that were in their BA

studies (without an undergraduate diploma) had more significant impact in their holistic assessment. Although, the analysis of the reasons behind this finding were out of the scope of this study, it is my belief that while teacher students may be learning to be more analytical and still need to practice this skill, their teacher peers, who have already obtained their degree, may have more experience analysing students´ work and decomposing students´ performance to understand it from a bottom-up perspective. Reliability analysis also revealed that those that had an academic degree were more consistent thus more reliable in their analytic assessment. These results may add to those found by Barkaoui (2011) in regard to the consistency with which those with an undergraduate or graduate degree assess in comparison to those that are still in their studies.

Participants who had an undergraduate degree were more consistent among their analytic assessment therefore converging with the results obtained in Barkaoui´s (2011) study in which twenty-nine experienced raters and thirty-one novice teachers scored analytically and holistically twenty-four essays. The researcher intended to describe how specific rubric variables and assessor variables had an important role in holistic assessment. Data revealed that those who were enrolled in a teaching program provided more varied and less predictable scores than those that had a BA and MA degree.

6.8 Chapter Summary

As mentioned in the previous sections, impact of WAT was identified in the affective and cognitive processes teachers experienced when confronted with a new experience. Evidence of this study suggested that a higher degree of awareness of the importance of

writing in the EFL classroom, increase in students´ awareness of the importance of the inclusion of writing in classroom assessment, and an increase in teachers´ reflection in regard to their performance as a teacher and as an EFL assessor were among the major gains of WAT. The results obtained allowed the construct of the Writing Assessment Training Impact Categorization (WATIC) which is a multi-level categorization that portrays the different kinds of impact that training may bring upon teachers.

In regard to teachers´, students´ and language managers´ perceptions of assessment it was found that they converged and diverged in different aspects. For instance, both language managers and teachers considered that two sessions of two-three hours each was not enough time to reflect on assessment and produce a change in their regular classroom assessment procedures or their assessment policies in their EFL programs. On the other hand, they also agreed that the biggest constraint to writing assessment was the lack of time teachers have in the classroom. Agreement was also found among teachers and students in the sense that they viewed teacher training as crucial to improve the assessment of the language and to raise awareness of the importance of assessing writing in the classroom. On the contrary, disagreement was found among teachers and students since the latter did not consider the former´s assessment procedures as fair and reliable since they felt they were not being assessed to help them learn but instead they were only being assessed to fulfil an institutional requirement. Finally, the three stakeholders´ changed their perspectives in regard to writing after teachers´ experienced WAT and reported to have increased the level of importance given to the learning and assessment of writing.

Results obtained from this study may also suggest that WAT may be a trigger to pursue future professional training. TP23, TP37 and TP325 explained during the interviews that they pretended to seek other courses that could help them improve their writing skills and their future assessment activities. Although the form of further development was not specified, experiencing WAT triggered their reflection regarding their strengths and weaknesses as a teacher and as an assessor. It awoke their desire to be more academically prepared since training may allow them to improve their professional practice. Therefore, it can be argued that teacher training may be an initial step that may trigger further assessment literacy and/or professional development.

Reliability analysis of teachers´ holistic scores suggested that consistency levels slightly increased post WAT while analytic assessment did not. However, these differences were not found to be of significant impact since excellent levels of reliability were found pre- and post to training. On the other hand, results pointed out that gender was not an impact factor on holistic scoring while analytic scoring did seem to be impacted by this trait. It can be argued that a holistic approach may allow more reliable assessment since it was found to be less affected by gender differences. Teacher teaching experience did not result to have an effect on the scoring of writing therefore suggesting that teachers who are novice to teaching EFL do not score more or less reliably than their more experienced peers. However, more needs to be done to explore the impact that years of experience in language assessment rather than EFL teaching may do to help or not the outcome of assessment. Finally, the analysis of the role of academic background in the scoring of the samples depicted an impact on the reliability of scores. It was surprising to find that those without an undergraduate degree scored more consistently both holistically and analytically.

Therefore, this finding may suggest that the scoring of language performance is influenced by the academic degree that the assessor holds.

The present study has attempted to provide an analysis of the factors that have been impacted by WAT. The following Chapter provides my conclusions of the results obtained and the possible implications that they may have.

Chapter 7: Conclusion

This Chapter presents some final ideas that emanate from the results obtained from this study. Firstly, some concluding remarks in regard to assessment training impact, teacher cognition, stakeholder perceptions and score reliability are discussed. Then, limitations observed during the development of the study are described followed by a description of possible contributions of this project to the field of language assessment. The Chapter then moves on to suggest future research ideas that may emerge from the limitations of this study. Finally, implications for EFL classroom assessment, the EFL curriculum as well as for teacher assessment literacy are presented in the last section of this Chapter.

7.1 Concluding remarks

The present study had the objective of analysing the extent to which WAT caused change in teachers´ reported classroom assessment of writing, stakeholders´ perceptions, and the analytic and holistic scores provided to five opinion essay samples. The results of this study suggest that WAT had effects on the encouragement of teacher cognition, on stakeholders´ perceptions of assessment, teacher assessment literacy and finally on the scores provided to students´ writing. The following sections focus on the description of my concluding interpretations of the results obtained.

7.1.1 Impact of Writing Assessment Training on EFL Assessment Stakeholders

Qualitative data collected during the study strongly suggested that the most notorious impact of WAT was on teachers´ and language managers´ awareness of the importance of including writing assessment in their EFL classroom for the development of students´

language skills. The evidence collected also suggested that WAT triggered participants´ self-awareness of their role as EFL teachers and/or writing assessors and their analysis of their own assessment processes in their classroom. Therefore, the WATIC (Figure 7, Chapter 6) was proposed with the intention of categorizing the distinct types of effects that training may encourage in EFL teachers. It is my belief that the WATIC may layout the impact of training to guide teacher trainers, language managers and heads of educational/language institutions in the understanding of the specific changes that may be encouraged in EFL teachers.

By understanding the possible changes, assessment stakeholders´ may have a better perspective of what needs to be included in training sessions as well as the extent to which teachers need to be trained. I also consider that by knowing what specific changes can be encouraged in teaching and assessment practice, financial resources can be more wisely allocated to teacher training.

7.1.2 Encouraging Teacher Cognition through Writing Assessment Training

It was surprising to discover that WAT did not have a significant impact on teachers´ actual classroom practices and the scores they provided. However, training was found to be a triggering component of teachers´ reflection processes. On this occasion, participating teachers and language managers reported that it was through the WAT that they identified their weaknesses and intended to pursue further development to improve their assessment performance. In other words, training allowed teachers to reflect on their practice as teachers, and as assessors subsequently leading to the planning of potential strategies to improve their practice (Sheehan and Munro, 2017). It can be argued that even small

amounts of time allocated to teacher training can encourage these reflections and trigger the improvement of assessment practice. For this study, two sessions of approximately two to three hours each was proven to trigger more impact on teachers´ self-awareness of a variety of different aspects related to assessment than any other aspect. Thus, my belief that even small amounts of time dedicated to teacher reflection may prove to be of great benefit.

It was also found that WAT gave opportunities to one of the language managers of analysing the actual assessment purposes and processes pursued by the EFL program she administered. Tangible innovations to these processes were only conducted to the nature of the writing assessment tasks included in their monthly tests which may seem a shallow innovation of low impact. However, the actual implementation of this change may suggest that WAT impacted at least one of the many processes that assessment entails and may leave a door open for future assessment innovation.

7.1.3 Student Perceptions and Classroom Writing Assessment

In relation to the EFL students participating in this study, results suggested that they considered writing and its assessment needed to be included in the English program for them to become better English users. It is my personal perception that these students were actually eager to develop their writing due to their comments during the focus group sessions and their interest in having daily writing activities conducted by the teacher. However, they also reported to feel afraid of its assessment due to the lack of its practice in their classroom. The minimum amount of contact students´ were having with writing in their regular EFL lessons, produced a sense of anxiety and discomfort with its assessment.

Data yielded from the collection and analysis indicated that once teachers experimented the training sessions, teachers changed their perception therefore having a domino effect on students´ views of the skill. In some cases, students reported that their teacher had increased the number of activities dedicated to writing in the classroom and that made them feel more comfortable with their learning. Therefore, this may support my claim that assessment training may not only lead to potential teacher improvement but may also have a positive impact on students: a) positive views towards writing, b) an increase of positive attitudes towards the practice of writing in the classroom, and c) a more comfortable acceptance of writing assessment criteria.

7.1.4 Impact of Training on Analytic and Holistic Scoring

The analytic and holistic scores that participating teachers gave to five student opinion essay samples revealed that holistic reliability was improved to a minimum post to training. It was also surprising to discover that the changes found were not significant enough to claim that WAT had an impact on teachers´ writing assessment.

When comparing analytic and holistic scores, holistic assessment was found to be more consistent in comparison to analytic scores. I believe that the fact that teachers had not previously used the specific scoring scale for this study (Appendix I) and the small amount of training sessions provided to teachers could have had an impact on the outcome obtained. Only two sessions were provided due to the time limits of this study during which the use of the analytic and holistic scoring tools was discussed and practiced. However, more sessions could have resulted in teachers´ improvement of assessment processes. On the other hand, other variables may have triggered these results, such as the

small number of samples teachers scored or the interpretation they may have given to each descriptor.

This study also analysed different teacher variables and its impact on assessment. Results indicated that teacher academic background had a significant impact on holistic (TPs without an undergraduate degree) and analytic assessment (TPs with a degree) thus suggesting that teachers who are more professionalized may portray more analytical abilities that may lead to higher levels of objectivity and reliability when using an analytic scoring tool. Results also suggested that holistic assessment may be a more reliable assessment approach since it was less influenced by teacher gender differences. In other words, the difference of holistic scores among women and men did not result in significant impact post to training.

This study was developed following specific procedures to ensure its objectivity and the credibility of its results. Ethical considerations were also considered in an attempt to protect the identity of participants. However, as any other study, the methodological design, the focus of the project and the results described in Chapter 5 may entail some limitations. These are further described in the following section.

7.2 Limitations of the study

While conducting this research project, many constraints were present and at times challenged the fulfilment of this study. However, several strategies were implemented to validate the research procedures followed so the results obtained would not be jeopardized. This section focuses on the description and justification of these limitations.

7.2.1 Methodological limitations

Sustaining TPs´, PMs´ and students´ commitment to participate in the study was a difficult issue that resulted in the limited number of participants and potential data loss (Hobbs and Kubanyiova, 2008). This project included the quantitative perceptions of forty-eight participants of which only eleven TPs participated to provide their qualitative views of the phenomenon under analysis. Thus, the construction of the WATIC included the views of only eleven active EFL teachers therefore more qualitative insight may provide a more valid and objective view of the effects of training. The number of language program managers and student focus groups was also limited therefore the perceptions provided by these participants portray those in this specific context only.

Approximately thirty to forty participants initially had agreed to participate in both training sessions but only took part in the first session of WAT therefore their participation was not considered in this study since they did not complete the two sessions. Although, qualitative insight may not pursue the generalization of knowledge (Dörnyei, 2007; Cohen, Manion, and Morrison, 2011) recruiting more participants in future research projects that can provide a wider perspective of their personal assessment literacy experiences and the impact of these on their regular assessment activities may allow the research community to gain a further understanding of the nature of assessment literacy in the Latin American context.

An additional issue added to the recruitment of participants was the intense commitment that TPs underwent throughout the study. The eleven participating teachers took part in five data collection phases (as outlined in section 4.7 of this thesis) that required their

investment of time and an additional workload to their already heavy workloads (Hobbs and Kubanyiova, 2008) without any type of financial incentive. The benefits of their participation such as academic development without cost were fully explained. But, in the Mexican context where teaching jobs are low-paid these benefits may not represent a significant gain. For others, such as the eleven participants who carried on participating in phases four and five, these benefits were enough. This intense workload experienced by the TPs may have resulted in data loss as well as participant withdrawal (Ibid).

Due to the nature of the Mexican EFL context, in which writing assessment is very frequently done in teachers´ out-of-classroom time, the impact of WAT was determined from the TPs voice instead of classroom observation. The researcher interviewed teachers prior and post to training and obtained the impact of training according to their views and their own experience. This may be considered an important limitation since the researcher was the trainer and the interviewer of the study. Thus, data obtained from teacher interviews and which led to the proposal of the WATIC (Figure 7) could be to some extent biased and influenced by TPs desire of performing how they believe the researcher expected them to perform (Dörnyei, 2007, p.53). To diminish this as much as possible strategies such as data triangulation methods, during which data was obtained from multiple data sources, was implemented to analyse information from multiple perspectives (as outlined in section 4.8.3 of this thesis).

Finally, it is important to point out that the time available throughout this research project as well as the financial commitments of being a PhD student supported by the Programa

para el Desarrollo Profesional Docente (PRODEP, Spanish acronym), or the Program for Teacher Professional Development (as translated to English) limited the proposal of the WATIC to its construction. In other words, its validation was not approached during this project. Validating the specific categories that attempt to describe the effects of WAT are crucial since this would allow the construction of valid and objective categories. It would be ideal to validate this categorization by exploring if with other teachers in similar and different contexts the same type of impact applies. This research idea is further explained in section 7.4 of this document.

7.2.2 Research focus limitations

With the intention of understanding the views of different stakeholders in relation to the assessment of writing and of assessment training, interviews were conducted with TPs and PMs. Students were also included in the study through focus group interviews. However, their views were less emphasized throughout the development of this project. Teachers´ and coordinators´ interviews were of longer length and more in-depth reflection was elicited from these stakeholders. Additionally, the WATIC was constructed considering only the views of the TPs. Further exploration of students´ views would allow the shaping of assessment literacy interventions to suit the needs of the EFL classroom.

As outlined in section 2.1, Flower and Hayes (1981) and later on Hayes (1996, 2012) proposed writing process models that attempt to portray the possible cognitive processes that L1 writers may experience. These Models may adjust to foreign language (FL) writers considering additional FL learning factors such as students´ linguistic knowledge, their intended audience, audience expectations, and their affective traits (such as motivation,

anxiety, self-awareness). On the other hand, these models portray different stages in the process of writing that classroom environments may not allow to be fully conducted. Guiding students´ through all the stages of the writing process, as portrayed in these models, may take large amounts of time that teachers may not have available in class. The consideration of these models as frameworks that portray the students´ process when writing in EFL may also be considered a limitation to this study since it does not genuinely represent students´ FL writing process in the Mexican context or teachers´ teaching of writing.

Consequently, important involved variables to these models need to be considered when adapting them to FL writing and its learning/teaching process. Factors such as students´ L1, their linguistic knowledge of the L2/FL, their knowledge of the genre, teachers´ teaching context (teaching time, language program), teaching tools available, teachers´ knowledge of writing, among others are important factors that need to be considered in FL writing models.

Another limitation of this study was the emphasis on the analysis of WAT and its effects on teachers without considering the actual content of these sessions. The sessions were adapted from the guidelines provided by CEFR Manual for Language Examinations (Council of Europe, 2002, 2009a, 2009b), the ALTE Manual for Language Test Development and Examining (Council of Europe, 2011) and the principles suggested by Bachman and Palmer (2010). However, if the content were considered for the analysis of WAT impact the results may have been different. Different training focus may lead to different outcomes and different reactions in teachers. Therefore, focusing on the various

effects that diverse types of training content may bring about in the EFL classroom could provide a wider perspective of the possible gains of assessment training.

A third limitation identified in regard to the training provided to TPs, was that only two days for each session were invested due to the limited availability and time that teachers had to actively participate in the study. Each session lasted approximately two to three hours, therefore the time available to engage in a deeper analysis of the assessment of teachers was limited. Participants of this study did not only work at the public university or the language institute under analysis, they also worked in other jobs or schools which kept them very occupied and with limited availability to attend sessions (Hobbs and Kubanyiova, 2008). However, it can be argued that small amounts of time may also lead to effects in teachers' conception or construction of their interpretation of language assessment and its application in the EFL classroom.

The main results of this project suggested that holistic assessment resulted more impacted post to training by TPs who did not have an undergraduate diploma while analytic assessment resulted more impacted by those that held an undergraduate degree. This study focused on identifying which specific teacher variables caused an impact on the scores provided to the opinion essays. However, the reasons behind these results were unexplored. It would be worthwhile to analyse these reasons so this phenomenon can be fully understood.

On the other hand, forty-eight participants scored analytically and holistically five sample papers. Therefore, each TP scored a total of ten sample papers analytically and ten holistically (five prior to training and five post to training). It is considered that the limited number of samples scored by TPs in this study may bias the results obtained since five samples could give scorers more room to remember previously assigned scores, or limited samples may provide different results when running the t-tests. Therefore, more sample papers to score is recommendable so results obtained from calculations have more possibilities of being generalized.

Although important limitations to this study have been identified, I believe this project has important contributions to the field of language assessment and assessment literacy which are pointed out in the following section.

7.3 Contributions to the Field of Language Assessment and Assessment Literacy

This study focused on the analysis of the effects of WAT on teachers´ classroom writing assessment. Language assessment literacy has been approached by researchers with a general perspective in which teachers´ needs, their perceived assessment literacy (Lam, 2015; Lopez Mendoza and Bernal Arandia, 2009; Vogt and Tsagari, 2014) or their perception of experienced assessment courses (Malone, 2013) have been analysed considering all the language skills. Approaching assessment literacy from a generalized perspective could cause assessors or researchers to lose focus. Therefore, I believe assessment literacy could be tackled by narrowing down its focus as has been done in this study. So, it can be considered that one of the main contributions of this study is the

provision of training impact on a focalised language skill: writing assessment. It would be worthwhile to analyse the impact of other skill training such as speaking assessment training to construct a categorization of each skill.

Another finding of this study was the substantial effect of WAT on teachers´ meta-cognitive skills. They became more aware of their need to improve their teaching of writing, improve their own writing skills, innovate their assessment procedures, increase their assessment of writing in the classroom and above all they became aware of their own perspective towards teaching and assessing writing. Few studies have set out to explore the benefits of training to teachers´ actual classroom assessment since teacher training has been considered the mostly used assessment literacy strategy (Hobbs and Kubanyiova, 2008). Those that have explored these benefits have provided a quantitative perspective (Barkaoi, 2011; Cheng *et al.*, 2004; Contreras, Gonzalez and Urias, 2009; Myford, 2013; Jin and Jie, 2017; Knoch, 2009; Shohamy *et al.*,1992; Weigle, 1994, 1998) in which teachers score writing samples post to training or their processes are quantified from their responses to surveys. This study has set out to understand WAT effects from a qualitative perspective to construct the proposed WATIC (Figure 7, Chapter 6). It has also set out to integrate the qualitative with the quantitative data obtained throughout the various collection phases. Therefore, it can be stated that one of the major contributions of this study to the literature regarding assessment literacy are the possible effects that training may bring upon teachers, their teaching practice and their assessment procedures as visualized in the WATIC. This categorization may allow teachers, teacher trainers and language managers understand the benefits that may be gained by experiencing WAT and

could allow them to make more accurate decisions as to which type of training is worth experiencing.

Other studies have set out to understand teachers´ knowledge of assessment, their assessment needs (Hasselgreen, Carlsen and Helness, 2004; Stiggins, 1999; Metler, 2003; Metler and Campbell, 2005; Fulcher, 2012; Tsagari and Vogt, 2017; Xu and Brown, 2017) or teachers´ procedures to assess foreign languages (Inbar-Lourie and Donitsa-Schmidt, 2009) from contexts such as those in Europe, the United States, England or Australia. Therefore, it can be stated that the Latin American context has been underexplored. This project attempts to contribute to the field of language assessment literacy by providing the Mexican perspective of how a tool to increase assessment literacy in teachers can actually help them in their assessment procedures. It provides an insight in regard to teachers´, program managers´ and students´ difficulties when teaching, learning and assessing writing in the Mexican EFL context. Above all, it provides the contextual factors such as, institutional assessment culture, program overload and time constraints in the classroom, that have a strong role in the Latin American context.

Other researchers have focused on teachers´ or language assessment experts´ perceptions (Vogt and Tsagari, 2014; Lopez Mendoza and Bernal Arandia, 2009; Nier *et al.,* 2013) of training sessions or of the usefulness of training to their assessment procedures. However, it is my belief that students´ perceptions have yet to be explored. This study, includes students´ views in regard to the importance of assessing writing, their teachers´ classroom assessment procedures and the importance of teacher assessment literacy. By understanding students´ views of teachers´ assessment literacy, or in this case, of the

assessment of writing, the teacher accounts for their opinion and gives it the importance it deserves since it is students who are one of the main beneficiaries of teacher improvement.

Analysis of assessors´ scores has been conducted considering their individual traits which may have an effect on the scores they provide to a text (Barkaoi, 2011; Lim, 2011; Myford, 2013; Wiseman, 2012). Results have suggested that analytical scoring has been the most reliable while those with the most experience may score more reliably students´ written performance. However, the contributions of this project point to the inclusion of academic background as a variable which was of significant impact on the reliability of scores. Additionally, the results of this study suggested that holistic scoring was less impacted by the gender of scorers. Therefore, contributing to the literature of writing assessment by pointing out holistic assessment as the most gender-free and reliable approach to assess EFL writing. This finding may provide possible insight to other researchers that may lead to the inclusion of this variable when conducting writing assessment research therefore leading to future research opportunities. These are further described below.

7.4 Opportunities for Future Research

As mentioned previously, this study proposed the creation of the Writing Assessment Training Impact Categorization (WATIC) which describes the potential innovations that WAT may encourage in EFL teachers and their assessment. It is my belief that future research is necessary to validate the WATIC with the intention of implementing any future adjustments that may be necessary as a result of its validation. This validation may be conducted with other EFL teachers in this country or other countries by observing their

training impact or trying to fit the impact to the categories in the WATIC. Validation may also include addition of missing categories, adaptation of existing categories or the proposal of a new WATIC.

I consider that it may be worthwhile to further investigate the implications that teacher training may have for the EFL student. In this study, teachers, language managers and students were interviewed to analyse their views in relation to writing assessment and assessment training. However, not much emphasis was placed on students. Future research may focus on analysing how teacher training can actually benefit students and their language development. In the long run, teacher training seeks to improve teacher classroom practices in the benefit of students. This focus would allow the development of an assessment categorization, such as the one proposed in this study (WATIC, Figure 7), that focuses specifically on students and their improvement.

Additionally, the focus of the present study was on the impact that WAT had on reported writing assessment practices without analysing the content that was approached during the training sessions. It is my belief that the content that is approached in assessment training or any type of training strongly influences the impact that it may cause. Thus, future research may seek to understand how specific WAT content influences teachers´ assessment performance. Content of training may vary greatly depending on the context or the attendants thus different approaches to the content of training may trigger distinct outcomes which may be worthwhile analysing. Two sessions of assessment due to time constraints and availability of teacher participants were provided to fulfil the purposes of this study. However, the influence of time and the number of sessions provided to teachers

was not analysed. Future research could explore if providing more sessions spread out over longer periods of time could actually benefit or hinder teachers´ development and/or improvement of their classroom assessment procedures. These future research ideas could allow teacher trainers or program managers to view how distinct amounts of time and content allocated to assessment training may produce effects in teachers to decide the amount of time and the type of content best suits their needs.

The present project focused on the practices that teacher participants reported to the researcher prior and post to training without considering classroom observation due to the nature of classroom assessment in the Mexican EFL context. In Mexico, writing assessment is very frequently done by the teacher during their free time. Most of the teachers are required to provide a specific score on a regular basis without the need of formative feedback that may allow students to improve. Therefore, observing a classroom seemed inconvenient to fulfil the main purpose of this study. Yet, it would be interesting to include the analysis of how teachers assess writing in the classroom and how this may or may not be connected with successful performance in a large-scale test, which in the Mexican context, is a requirement to obtain an undergraduate degree (Metler and Campbell, 2005).

I consider that the results obtained from the extensive amount of research conducted in the field of language assessment in addition to the results obtained in this study could have positive implications for EFL teachers and their classroom practices. Taking into account the results of the present study, I sought to put forward these implications for teachers in the north-eastern part of Mexico.

7.5 Implications for EFL Instruction and Assessment

Some of the main issues that were evident from stakeholders´ participation in this study were the little amount of time teachers have to teach and assess writing in their classrooms, the lack of importance that EFL programs give to the skill and its assessment, the lack of institutional support to innovate assessment procedures and provide training opportunities to teaching staff, and professors´ unreliable use of scoring rubrics. Therefore, these issues have strong implications for EFL instruction and/or assessment thus I consider specific measures can be taken to tackle negative washback.

7.5.1 Implications for the EFL Curriculum and the EFL Classroom

Results of this study suggest that the biggest constraint to the teaching and assessment of writing in the Mexican EFL context is the lack of time teachers have in the classroom to dedicate to the skill. In this case, it can be suggested to program managers to analyse jointly with teaching staff the contents of the EFL program that are being taught to students so that if content approached gives priority to other linguistic aspects and leaves aside the development of communicative language skills, then priorities need to be changed to allow more time to be spent on writing.

Some teacher participants suggested the creation of a writing or speaking club in which students dedicate solely to the development of language production skills. This would allow teachers more time to develop the skill with their students additionally to providing them with opportunities to improve their teaching and assessment practice. This EFL program innovation would allow the participation of managers, teachers and preferably students therefore promoting awareness of the need to teach and assess writing.

The results of this study suggest that standardization of context specific assessment procedures may enlighten the path to valid and reliable assessment. Teachers and language managers can take part in this standardization process. For instance, teachers who teach a common level of English proficiency may agree on specific assessment criteria that can be updated depending on the needs of the students. Additionally, these stakeholders can also adapt or create scoring tools so that assessment of writing (and possibly speaking abilities) can be more objective and easier for EFL classroom teachers´ practice. Students may be involved in the standardization process by encouraging their participation in the adaptation or creation of scoring tools by eliciting their opinion or needs in terms of the skill assessed or their views in terms of the criteria that should be (according to their opinion) included in their regular skill assessment.

Finally, it is my belief that the difference between summative and formative assessment needs to be analysed in terms of the boundaries of each concept. This may lead to the creation or implementation of a third approach to assessment that involves a combination of both types of assessment in a language classroom. For instance, in the Mexican context as in many other parts of the world, teachers perform summative and formative assessment of their students to comply with the institutional/administrative requirements. Thus, considering summative assessment not only in large-scale testing contexts but also in classroom contexts where teachers score their students´ performance without providing feedback. This combination of assessment approaches conducted in practice by language teachers may portray a more pragmatic stance towards assessment implying the need to consider the role of small scale summative assessment.

7.5.2 Implications for Teacher Assessment Literacy

It is important to consider the benefits of WAT. However, for these benefits to be tangible in practice, teachers need to have sufficient time to assimilate and reflect on the contents discussed during sessions. I believe that since teacher self-reflection is a complex process that may depend on many variables out of reach to the researcher, EFL teachers need to be provided with not only one training session, but multiple sessions that can on several occasions provide them with opportunities to analyse their present assessment context and potential ways of improving it.

It may also be worthwhile analysing teachers´ needs and perceptions prior to the implementation of assessment training sessions. By understanding these, content of training sessions can be structured, contextualized and suited to the specific institution, the institutional culture and potential teachers´ needs. This may allow, training sessions to have a higher degree of impact in language instructors´ classroom assessment.

The WATIC (Figure 7, Chapter 6) may also be a tool for teacher trainers to predict the potential effects their training may cause. These potential effects may allow trainers to plan ahead the contents of their workshops to correspond the desired effects. Institutions may also find the results of this study useful considering that they need to provide their staff with constant and permanent training opportunities that may be organized financially and academically in accordance to the desired impact on teachers´ assessment. These previously described measures may seem complicated. However, the gains that teachers will have in their professional practice and the potential benefits that students will have in their learning may well be worth the effort. It is my belief that by considering the

importance that assessment training has for teachers, the language program and the language institution, teachers are valued for their crucial role in the success of an institution. It is also a way that teachers, managers and institutions value the English language teaching profession in benefit of language students because `after all the whole purpose of education is to turn mirrors into windows ´ (Sydney J. Harris, n.d.).

List of References

Abdul Raof, A.H. (2002). *The Production of a Performance Rating Scale; An Alternative Methodology*. Unpublished PhD thesis, The University of Reading.

Abdul Raof, A.H.; Hamzah, M.; Aziz, A.A; Mohd, N. A.; Atan, O. and Atan, A. (2011). Profiling graduating students´ workplace oral communicative competence. IN Powell-Davies, P. (ed.) *New Directions: Assessment and Evaluation A collection of papers*, Scotland: British Council, 155-160.

Assessment Reform Group (2002) *Assessment for Learning: 10 Principles.* United Kingdom: Nuffield Foundation.

Attali, Y. (2015) A comparison of newly-trained and experienced raters on a standardized writing assessment. *Language Testing*, 33 (1), 99-115.

Atan, A.; Abdul Raof, A.H.; Mohammed Yusof, M.A.; Mohammed Omar, N.A. and Hamzah, M. (2015) Determining the oral construct of the Test of English Communication Skills. *International Journal of Economics and Financial Issues*, 5 (1S), 139-143.

Bachman, L. (2004) *Statistical Analyses for Language Assessment*. Cambridge, United Kingdom: Cambridge University Press.

Bachman, L.F. and Palmer, A. (2010) *Language Assessment in Practice*. Oxford, U.K.: Oxford University Press.

Bailey, K. M., and Brown, J. D. (1996) Language testing courses: What are they? IN: Cumming, A. and Berwick, R. (eds.) *Validation in language testing*, London, UK: Multilingual Matters, 236–256.

Barkaoui, K. (2007) Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing*, 12(2), 86–107.

Barkaoui, K. (2011) Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy & Practice,* 18(3), 279–293.

Black, P. and William, D. (1998a) *Inside the Black Box: Raising Standards through Classroom Assessment*. London: School of Education, King's College.

Black, P. and William, D. (1998b) Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5 (1), 7-74.

Brindley, G. (2001) Outcomes-based assessment in practice: some examples and emerging insights. *Language Testing*, 18 (4), 393–407.

Brown, H.D. (2007) *Teaching by Principles: An Interactive Approach to Language Pedagogy*, 3rd ed. Harlow, United Kingdom: Pearson Longman.

Brown, H.D. and Abeywickrama, P. (2010) *Language Assessment: Principles and Classroom Practices,* 2nd ed. Harlow, United Kingdom: Pearson Longman.

Brown, A. (2012) Ethics in Language Testing and Assessment. IN: Coombe, C.; Davidson, P.; O'Sullivan, B. and Stoynoff, S. (eds.) *The Cambridge Guide to Second Language Assessment*, New York, NY USA: Cambridge University Press, 113-132.

Byrne, D. (1991) *Teaching Writing Skills*. Essex, England: Longman.

Charmaz, K. (2008) Grounded Theory as an emergent method. IN S.N. Hesses-Biber and P. Leavy (eds.), *The Handbook of Emergent Methods*, New York, USA: The Guilford Press, 155-172.

Chen, Q., Kettle, M.A., Klenowski, V., and May, L. (2013) Interpretations of formative assessment in the teaching of English at two Chinese universities: a sociocultural perspective. *Assessment and Evaluation in Higher Education*, 38(7), 831-846.

Cheng, L., Rodgers, T. and Hu, H. (2004) ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. *Language Testing*, 21 (3), 360–389.

Cheng, L.; Rogers, W.T. and Wang, X. (2008) Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, 33, (1), 9-32.

Cheng, L. and Wang, X. (2007). Grading, Feedback, and Reporting in ESL/EFL Classrooms. *Language Assessment Quarterly*, 4(1), 85–107

Cheng, Yuh-show, Horwitz E.K. and Schallert, D.L. (1999) Language Anxiety: Differentiating Writing and Speaking Components. Language Learning, (49)3, 417-446.

Cicchetti, D.V. (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284.

Cohen, L., Manion, L. and Morrison, K. (2011) *Research Methods in Education*, 7th ed. Oxford, UK: Routledge.

Contreras Niño, L.A., González Montesinos, M. and Urías Luzanilla, E. (2009) Evaluación de la escritura mediante rúbrica en la educación primaria en México. *Interamerican Journal of Psychology*, 43(3), 518–531.

Coombe, C., Troudi, S. and Al-Hamly, M. (2012) Foreign and Second Language teacher assessment literacy: Issues, challenges and recommendations IN: Coombe, C., Davidson, P., O'Sullivan, B. and Stoynoff, S. (eds.) *The Cambridge Guide to Second Language Assessment*, New York, NY USA: Cambridge University Press, 20-29.

Coordinación de Inglés en Educación Básica (2015) Antes de CIEB Coordinación de Inglés en Educación Básica. Available at: http://www.programa-ingles.net/site/index.php?url=about.php&menu_id=2 [Accessed February 2015].

Council of Europe. (2002) *Common European Framework of Reference for Languages: Learning, Teaching and Assessment*. Strasbourgh, France: Council of Europe.

Council of Europe. (2009a) *The Manual for Language Test Development and Examination*. Strasbourgh, France: Council of Europe.

Council of Europe. (2009b) *Manual for Relating Language examinations to the Common European Framework of Reference for Languages: Learning, Teaching and Assessment*. Strasbourgh, FR: Council of Europe.

Council of Europe. (2011) *Manual for Language Test Development and Examining: For use with the CEFR.* Strasbourgh, France: Council of Europe.

Creswell, J. W. and Plano Clark, V. L. (2011) *Designing and Conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.

Creswell, J. W. (2013) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, US: Sage Publications.

Creswell, J.W. (2015) *A Concise Introduction to Mixed Methods Research.* Thousand Oaks, California, USA: Sage Publications.

Crossman, J. (2007) The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research & Development*, 26 (3), 313-327.

Crusan, D. (2010) *Assessment in the Second Language Writing Classroom*. Ann Arbor, Michigan, USA: The University of Michigan Press.

Crusan, D. (2014) Assessing Writing IN: Kunan, A.J. (ed.) *The Companion to Language Assessment*, West Sussex, UK: John Wiley & Sons, 206-217.

Cumming, A. (2001) ESL/EFL instructors' practices for writing assessment: specific purposes or general purposes? *Language Testing*, 18(2), 207-224.

Daniels, P. (2001) Writing Systems IN: Aronoff, M. and Rees-Miller, J.(eds.), *The Handbook of Linguistics,* Padstow, Cornwall United Kingdom: Blackwell Publishing, 43-80.

Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C., and BivensTatum, J. (2008). *Cognitive models of writing: Writing proficiency as a complex integrated skill (Research Report No. RR-08 –55)*. Princeton, NJ: Educational Testing Service.

Donald, J.G. and Denison, D.B. (2001) Quality Assessment of University Students: Student Perceptions of Quality Criteria. *The Journal of Higher Education*, 72 (4), 478-502.

Dörnyei, Z. (2003) *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z. (2007) *Research Methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies.* Oxford, United Kingdom: Oxford University Press.

Dörnyei, Z., and Taguchi, T. (2010) *Questionnaires in Second Language Research: Construction, Administration and Processing*, 2nd ed. New York, US: Routledge.

Dunne, R.A. (2007) The Exaver Project: Conception and Development in Mexico. *MEXTESOL Journal*, 31 (2), 23-30.

Educational Testing Service (2004) IBT/Next Generation TOEFL Test: Integrated Writing Rubrics (Scoring Standards). Available at

https://www.ets.org/Media/Tests/TOEFL/pdf/Writing_Rubrics.pdf Retrieved [April 12, 2017].

Elder, C., Knoch, U., Barkhuizen, G., and Randow, J. (2005) Individual Feedback to Enhance Rater Training: Does It Work? *Language Assessment Quarterly*, 2(3), 175-196.

Elder, C., Barkhuizen, G., Knoch, U., Randow, J. (2007) Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing*, 24(1), 37–64.

Esfandiari, R. and Myford, C. (2013) Severity differences among self-assessors, peer-assessors, and teacher assessors rating EFL essays. Assessing Writing, 18(2), 111–131.

Educational Testing Service (2017) About the TOEFL ITP Assessment Series. Available at https://www.ets.org/toefl_itp/about [Accessed August 30, 2017].

Ferris, D.R. and Hedgecock, J.S. (2014) *Teaching L2 Composition: Purpose, Process and Practice*, 3rd ed. New York, USA: Routledge.

Flower, L. and Hayes, J. R. (1981) A Cognitive Theory Process of Writing. *College Composition and Communication*, 32 (4), 365-387.

Fulcher, G. (2012) Assessment Literacy for the Language Classroom, *Language Assessment Quarterly*, (9)2, 113-132.

Fulcher, G. (2014) The Multiple-Choice test: Truly objective? Available at http://www2.le.ac.uk/projects/social-worlds/all-articles/education/multiple-choice [Accessed August 2, 2017].

Fulcher, G. and Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book.* Abingdon, Ox: Taylor and Francis.

Froetscher, D. (2017) Washback on classroom testing: Assessment literacy as a mediating factor. Paper presented at the 39th Language Testing Research Colloquium, Bogotá Colombia, July 17-21.

Gardner, R.C., Masgoret, A. and Tremblay, P.F. (1999) Home Background Characteristics and Second Language Learning. *Journal of Language and Social Psychology*, 18 (4), 419-437.

Glówka, D. (2011) Mix? Yes, but how? Mixed Methods Research Illustrated IN: Pawlak, M. (ed.), *Extending the Boundaries of Research on Second Language Learning and Teaching*, Poland: Springer, 289-300.

González, E.F. and Vega López N.A. (2018) Exploring Mexican EFL Elementary School Teachers' Perceptions of Online Language Assessment Training, 5 (1), 225-241. Available at http://www.jallr.com/index.php/JALLR/article/view/762/pdf762

Grabe, W. and Kaplan, R.B. (1996) *Theory and Practice of Writing*. Essex, England: Pearson Longman.

Hamp-Lyons, L. (1990) Second Language Writing: Assessment Issues IN: Kroll, B.(ed.) *Second Language Writing: Research Insights for the Classroom* Cambridge, U.K.: Cambridge University Press, 69-87.

Hamp-Lyons, L. (2001) Fourth Generation Writing Assessment IN: Silva, T.J. and Matsuda, P.K. (eds.) *On Second Language Writing*, Mahwah, New Jersey USA: Lawrence Erlbaum Associates, 117-128.

Hamp-Lyons, L. (2003) Writing teachers as assessors of writing IN: Kroll, B. (ed.) *Exploring the Dynamics of Second Language Writing*, New York, USA: Cambridge University Press, 162-189.

Harmer, J. (2007) *The Practice of English Language Teaching*. Harlow, England: Pearson Longman.

Harris, S. J. (n.d.) BrainyQuote.com. Available from: BrainyQuote.com Web site https://www.brainyquote.com/quotes/quotes/s/sydneyjha104885.html [Retrieved April 12, 2017]

Hasselgreen, A., Carlsen, C., and Helness, H. (2004) European Survey of Language Testing and Assessment Needs. Part 1: General findings. Gothenburg, Sweden: European Association for Language Testing and Assessment. Available at http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf [Retrieved April, 2014]

Hayes, J. R. (1996) A new framework for understanding cognition and affect in writing. IN: Levy, C.M. and Ransdell, S. (eds.) *The science of writing. Theories, methods, individual differences and applications,* Mahwah, New Jersey, USA: L.E.A, 1- 27.

Hayes, J.R. (2006) New Directions in Writing Theory IN: MacArthur, C.A.; Graham, S and Fitzgerald, J. (eds) *Handbook of Writing Research*, 1st ed. New York, USA: Guilford Publications, 8-40.

Hayes, J.R. (2012) Modeling and Remodeling Writing. *Written Communication*, 29(3), 369-388.

Hobbs, V. and Kubanyiova, M. (2008) The challenges of researching language teachers: What research manuals don´t tell us. *Language Teaching Research*, 12 (4), 495-513.

Huot, B. (2002) *Rearticulating Writing Assessment for Teaching and Learning.* Logan, Utah, USA: Utah State University Press.

Hyland, K. (2004) *Second Language Writing.* Cambridge, U.K.: Cambridge University Press.

Hyland. K. (2015) *Teaching and Researching Writing*, 3rd ed. New York, USA: Routledge.

Inbar-Lourie, O. (2008) Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25 (3), 385 – 402.

Inbar-Lourie, O. (2013) Guest Editorial to the special issue on language assessment literacy. *Language Testing*, 30 (3), 301 – 307.

Inbar-Lourie, O. (2017) Language Assessment Literacies and language testing community: A mid-life identity crisis? Lecture presented at the 39th Language Testing Research Colloquium, Bogotá Colombia, July 17-21.

Inbar-Lourie, O. and Donitsa-Schmidt, S. (2009) Exploring classroom assessment practices: the case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice,* 16(2), 185–204.

Instituto Nacional para le Evaluación de la Educación (2013) Acerca del INEE. Available at http://www.inee.edu.mx/index.php/acerca-del-inee [Accessed August 15, 2017].

Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., and Hughey, J. (1981) *Testing ESL composition: A practical approach*. Rowley, USA: Newbury House.

Jeong, H. (2013) Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, (30) 3, 345-362.

Jin, Y. and Jie, W. (2017) Do workshops really work? Evaluating the effectiveness of training in language assessment literacy. Paper presented at the 39th Language Testing Research Colloquium, Bogotá Colombia, July 17-21.

Jin, Y., Jin, T., Benigno, V., Jong, J., Gu, L., Yao, L., Davis, L., Zhu, B., Wang, W., Li, B. and Xi, X. (2017) Human-machine teaming up for language assessment: The need for

extending the scope of assessment literacy. Symposia presented at the 39th Language Testing Research Colloquium, Bogotá Colombia, July 17-21.

Johnson, R.B. and Onwuegbuzie, A.J. (2004) Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33 (7), 14-26.

Johnson, R.B., Onwuegbuzie, A.J. and Turner, L.A. (2007) Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2), 112-133.

Ketabi, S. and Ketabi, S. (2014) Classroom and Formative Assessment in Second/Foreign Language Teaching and Learning. *Theory and Practice in Language Studies*, 4, (2), 435-440.

Kitano, K. (2001) Anxiety in the College Japanese Language Classroom. *The Modern Language Journal*, 85 (4), 549-566.

Klenowski, V. (2009) Assessment for learning revisited: an Asia-Pacific perspective. *Assessment in Education: Principles, Policy and Practice*, 16 (3), 263-268.

Knoch, U. (2009) Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275–304.

Knoch, U., (2011) Investigating the effectiveness of individualized feedback to rating behavior -- a longitudinal study. *Language Testing*, 28(2), 179–200.

Koh, K., Burke, L.E.C., Luke, A., Gong, W. and Tan, C. (2017) Developing Assessment literacy in teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 1-25.

Kroll, B. (1991) Understanding TOEFL's Test of Written English. *RELC Journal*, 22 (1), 20 – 33.

Kroll, B. (1998) Assessing Writing Abilities. *Annual Review of Applied Linguistics*, 18, 219-240.

Lam, R. (2015) Language assessment training in Hong Kong: implications for language assessment literacy. *Language Testing*, 32 (2), 169-197.

Lee, I. (2007) Assessment for Learning: Integrating, Assessment, Teaching, and Learning in the ESL/EFL Writing Classroom. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 64 (1), 199–214.

Leech, N.L.; Barrette, K.C. and Morgan, G.A. (2014) *IBM SPSS for intermediate Statistics: Use and Interpretation* (5th ed). New York, USA: Routledge.

Leung, C. and Mohan, B. (2004) Teacher formative assessment and talk in classroom contexts: assessment as discourse and assessment of discourse. *Language Testing*, 21 (3), 335-359.

Lim, G. (2011) The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28, 543-560.

Lingard, L.; Albert, M. and Levinson, W. (2008) Grounded theory, mixed methods and action research. *BMJ*, 459-461.

Lopez Mendoza, A.A. and Bernal Arandia, R. (2009) Language Testing in Colombia: A Call for More Teacher Education and Teacher Training in Language Assessment. *PROFILE,* 11 (2), 55-70.

Mackey, A. and Gass, S. M. (2005) *Second Language Research: Methodology and Design*, Mahwah, New Jersey, USA: Routledge.

Malone, M.E. (2013) The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30 (3), 329–344.

Metler, C. (2003) Preservice versus In-service Teachers´ Assessment Literacy: Does Classroom Experience Make a Difference? Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, Ohio, 15-18 October.

Metler, C. and Campell, C. (2005) Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the Assessment Literacy Inventory. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, QC.

Moss, P. A. (1996) Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25(1), 20–28.

Moss, P. A. (2004) The meaning and consequences of reliability. *Journal of Educational and Behavioural Statistics*, 29, 241–245.

Nier, V.C.; Donovan, A.E. and Malone, M.E. (2013, October) *Promoting Assessment Literacy for Language Instructors through an online course*. Poster presented at the East Coast Organization of Language Testers Conference, Washington, D.C.

Nunan, D. (1992) *Research Methods in Language Learning*, New York, USA: Cambridge University Press.

O'Loughlin, K. (2013) Developing the assessment literacy of university proficiency test users. *Language Testing*, 30 (3), 363–380.

O´Malley, J.M. and Pierce, L.V. (1996) *Authentic Assessment for English Language Learners: Practical Approaches for Teachers.* Boston, USA: Addison-Wesley Publishing Company

Onwuegbuzie, A.J. and Leech, N.L. (2005) On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies. *International Journal of Social Research Methodology,* 8 (5), 375–387.

O´Sullivan, B. (2012) A Brief History of Language Testing. IN: Coombe, C., Davidson, P., O'Sullivan, B. and Stoynoff, S. (eds.) *The Cambridge Guide to Second Language Assessment*, New York, NY USA: Cambridge University Press, 9-19.

Pavlenko, A. (2007) Autobiographic Narratives as Data in Applied Linguistics. *Applied Linguistics*, 28(2), 166-188.

Pearson, P.C. (2004) *Controversies in Second Language Writing: Dilemmas and Decisions in Research and Instruction*, Michigan, United States: The University of Michigan Press.

Perry, F. (2011) *Research in Applied Linguistics: Becoming a discerning consumer,* 2nd ed. New York, New York USA: Routledge.

Phillips, D. (2009) *Longman Preparation Course for the TOEFL Test: The Paper Test,* White Plains, New York, USA: Pearson Education.

Polio, C. and Williams, J. (2011) Teaching and Testing Writing IN: Long, M.H. and Doughty, C.J. (eds) *The Handbook of Language Teaching*, 1st ed., West Sussex, U.K.: Blackwell Publishing, 486-517.

Randall, L. and Zundel, P. (2012) Students´ Perceptions of the Effectiveness of Assessment Feedback as a Learning Tool in an Introductory Problem-solving Course. *The Canadian Journal for the Scholarship of Teaching and Learning*, 3 (1), 1-16.

Rea-Dickins, P. and Gardner, S. (2000) Snares and silver bullets: disentangling the construct of formative assessment, *Language Testing,*17(2), 215-243.

Roever, C. and Phakiti, A. (2017) *Quantitative Methods for Second Language Research: A Problem-Solving Approach*. New York, USA: Routledge.

Roux, R. and Valladares Mendoza, J. (2014) *El Desarrollo Profesional Continuo de los Docentes: Teoría, Investigación y Practica*, Ciudad Victoria, Tamaulipas México: El Colegio de Tamaulipas.

Sambell, K.; Mcdowell, L. and Brown, S. (1997) "But is it fair?": An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23 (4), 349-371.

Sasaki, M. (2000) Toward and Empirical Model of EFL Writing Processes. *Journal of Second Language Writing*, 9, 259-292.

Saxton, E.; Belanger, S. and Becker, W. (2012) Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing*,17, 251–270.

Scarino, A. (2013) Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30 (3), 309–327.

Scarino, A. (2017) Developing Assessment Literacy of teachers of languages: A conceptual and interpretative challenge. Papers in Language Testing and Assessment, 6 (1), 18-40.

Sheehan, S. And Munro, S. (2017) Assessment: attitudes, practices and needs. *ELT Research Papers*. London, U.K: British Council.

Shohamy, E.; Gordon, C.M. and Kraemer, R. (1992) The Effect of Raters' Background and Training on the Reliability of Direct Writing Tests. *The Modern Language Journal*, 76 (1), 27-33

Shohamy, E., Inbar-Lourie, O., and Poehner, M.E. (2008) *Investigating assessment perceptions and practices in the advanced foreign language classroom* (Report No. 1108) University Park, PA: Center for Advanced Language Proficiency Education and Research.

Stiggins, R. J. (1995) Assessment literacy for the 21st century. *Phi Delta Kappan,* 77(3), 238-245.

Stiggins, R. J. (1999) Evaluating Classroom Assessment Training in Teacher Education Programs. *Educational Practice: Issues and Practice*, 18 (1), 23-27.

Stoynoff, S. and Coombe, C. (2012) Professional Development in Language Assessment IN: Coombe, C., Davidson, P., O' Sullivan, B. and Stoynoff, S. (eds.) *The Cambridge Guide to Second Language Assessment*, New York, NY USA: Cambridge University Press, 122-130.

Strauss, A. and Corbin, J. (1994) Grounded Theory Methodology: An overview. IN Denzin, N.K. and Lincoln, Y.S. (eds.) *Handbook of qualitative research*, Thousand Oaks, California, USA: Sage, 273-285.

Taber, K. S. (2000) Case studies and generalizability: grounded theory and research in science education. *International Journal of Science Education*, 22(5), 469- 487.

Tsagari, D. and Vogt, K. (2017) Assessment Literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.

Taylor, L. (2009) Developing Assessment Literacy. *Annual Review of Applied Linguistics*, 29, 21-36.

Taylor, L. (2012) Developing Assessment Literacy. Available from http://lrweb.beds.ac.uk/_data/assets/pdf_file/0010/197641/PROSET-Assessment-Literacy-Feb-212.pdf [Accessed February 13 2017].

Taylor-Powell, E. And Renner, M. (2000) Collecting Evaluation Data: End of Session Questionnaires. Program Development and Evaluation, University of Wisconsin-Extension: Madison, Wisconsin.

Teddlie, C. and Tashakkori, A. (2006) A general typology of research designs featuring mixed methods. *Research in the Schools*, 13(1), 12-28.

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.

Vogt, K. and Tsagari, D. (2014) Assessment Literacy for Foreign Language Teachers: Findings of a European University. *Language Assessment Quarterly*, 11(4), 374-402.

Volante, L. and Fazio, X. (2007) Exploring Teacher Candidates´ Assessment Literacy: Implications for Teacher Education Reform and Professional Development. *Canadian Journal of Education*, 30 (3), 749-770.

Wang, W. and Wen, Q. (2002) L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, 11, 225-246.

Wang, L. and Yan, X. (2017) Working towards professional standards for EFL test developers in China: An investigation into stakeholders´ perceptions of language testing practice. Paper presented at the 39th Language Testing Research Colloquium, Bogotá Colombia, July 17-21.

Weir, C.J. (1990) *Communicative Language Testing*. NJ, US: Prentice Hall Regents.

Weigle, S.C. (1994) Effects of Training on Raters of ESL compositions. *Language Testing*, 11, 97-223.

Weigle, S.C. (1998) Using FACETS to model rater training effects. *Language Testing*, 15, 263-287.

Weigle, S.C. (2002) *Assessing writing*. Cambridge, UK: Cambridge University Press.

Weigle, S.C. (2007) Teaching writing teachers about assessment. *Journal of Second Language Writing*, 16 (3), 194–209.

White, E.M., (1990) Language and Reality in Writing Assessment. *College Composition and Communication*, 41(2), 187–200.

Wiseman, C.S. (2012) Rater effects: Ego engagement in rater decision-making. *Assessing Writing*, 17 (3), 150–173.

Woodrow, L. (2014) *Writing about Quantitative Research in Applied Linguistics*. London, United Kingdom: Springer.

Xu, Y. and Brown, G. T. L. (2016) Teacher Assessment Literacy in Practice: A Reconceptualization. *Teaching and Teacher Education*, *58*, 149-162.

Xu, Y. and Brown, G. T. L. (2017) University English teacher assessment literacy: A survey test-report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158.

Yan, X., Fan, J. and Zhang, C. (2017) Understanding language assessment literacy profiles of different stakeholder groups in China: The importance of contextual and experiential factors. Paper presented at the 39th Language Testing Research Colloquium, Bogotá Colombia, July 17-21.

Yorke, M. (2003) Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education,* 45, 477–501.

Zamel, V. (1983) The composing Processes of Advanced ESL Students: Six Case Studies. *TESOL Quarterly*, 17 (2), 165-187

Appendices

Appendix A Teacher Background Questionnaire

This questionnaire is part of a research project that seeks to analyse the assessment strategies of EFL writing in distinct public universities of Ciudad Victoria Tamaulipas Mexico. This instrument has the purpose of finding out more about you and your experience evaluating strategies of the written ability. Please be so kind and honestly answer the following questions. Check the option that most suits your opinion or experience. There is no wrong or correct answer, only experiences to share. The information you share on this questionnaire is anonymous and confidential. It will only be used for research purposes. If you need any assistance with this questionnaire or have any questions regarding your participation in this research project feel free to contact the researcher via email: e.fernandagonzalez@gmail.com.

Participant ID: _____ Age: _____ Sex: a) M ☐ b) F ☐
Faculty at which you work: _____
Time of EFL teaching experience: _____
Academic preparation: _____

1. I evaluate my students´ writing as part of their academic progress throughout the course.
   ☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.

2. I use evaluation tools such as scoring rubrics to evaluate my students´ writing.
   ☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.
3. I consider that using a scoring rubric makes it easier to differentiate among students´ levels of writing.
   ☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.
4. I consider that using a scoring rubric makes my evaluation of written texts more objective.
   ☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.
5. I consider that using a scoring rubric makes my evaluation of written texts more efficient.
   ☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.
6. When evaluating my students´ text, I read the text several times and give several scores to different aspects.
☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.

7. When evaluating my students´ text, I read the text once and give it a single general score.

   ☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.

8. When I am not sure about the paper I scored, I ask a colleague or friend for their opinion about the text that is being evaluated.

   ☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.

9. When I am not sure about the score I gave, I ask a colleague or friend for their opinion.

   ☐ Always ☐ Often ☐ Sometimes ☐ Rarely
   ☐ Hardly ever ☐ Never.

10. Before today´s session, I have received specific training on the evaluation of writing. If yes, please describe this experience. If no, do you think it could improve your usual evaluation activities? Please comment on your answer.

    ☐ Yes ☐ No

11. Before today´s session, I have participated in teacher seminars or workshops that address the use of rubrics and other scoring tools. If yes, please describe this experience. If no, do you think it could improve your usual evaluation activities? Please comment on your answer.

    ☐ Yes ☐ No

12. Please give any additional comments you consider necessary to describe your current practice of evaluating writing or your usual marking process.

Thank you very much for your valuable information.

Appendix B Student Background Questionnaire

This questionnaire is part of a research project that seeks to analyse the assessment strategies of EFL writing in distinct public universities of Ciudad Victoria Tamaulipas Mexico. This instrument has the purpose of finding out more about you and your experience as an English learner. Please be so kind and honestly answer the following questions. Check the option that most suits your opinion or experience. There is no wrong or correct answer, only experiences to share. The information you share on this questionnaire is anonymous and confidential. It will only be used for research purposes. If you need any assistance with this questionnaire or have any questions regarding your participation in this research project feel free to contact the researcher via email: e.fernandagonzalez@gmail.com.

Participant ID: _____ Age: _____ Sex: a) M ☐ b) F ☐ Level of English of Study:_____
Place of Study:_____
Years studying English: _____

Have you take an English Language Examination before? If yes, please include which exam and score obtained. _____

Do you study an additional language? ☐ Yes ☐ No
If your answer is yes, please indicate which one_____

1. Do you like to study English? ☐ Yes Why?_____

☐ No Why?_____

2. Which is your main difficulty when writing in English?_____
_____

3. Does your English teacher require you to write in English? ☐Yes ☐ No

4. If your answer to question 3 is yes, answer the following:
   a) I write in my English class.

☐ Always ☐ Often ☐ Sometimes ☐ Rarely ☐ Hardley ever ☐ Never.

   b) Whay types of texts do you write in your English class? More than one option is valid. If you choose other please specify which text.

☐ Essays ☐ Letters ☐ Reports ☐ Journals☐ Descriptions of personal experiences

☐ Other:_____

c) Your teachers consider your written texts as part of your English grade.

☐ Always ☐ Often ☐ Sometimes ☐ Rarely ☐ Hardley ever ☐ Never.

d) Writing is an important part of my English evaluation.

☐ Strongly Agree ☐ Disagree
☐ Agree ☐ Strongly Disagree
☐ Neither Agree or Disagree

e) What percentage of your total English grade is given to writing?

☐ 5%-20% ☐ 25%-40% ☐ 45%-60% ☐ 65%-80% ☐ 85%-100%

6. If your answer to question 3 was NO, please answer the following:

a) Why do you consider writing is not considered an important part of your grade?

b) Would you like writing to be considered in your English grade?

1) ☐ Yes Why?_____

2) ☐ No Why?_____

c) Do you consider it important for the evaluation of writing to be considered in an English program?

1) ☐ Yes Why?_____

2) ☐ No Why?_____

Do you have an additional comment you would like to add?

Appendix C Teacher Interview 1 Outline

Questions about the teaching and assessment of writing.

1. Do you consider writing an important skill to develop in a language student? Why?
2. Do you teach writing in your classroom? How regularly? Do you consider it as a part of students' bimonthly or semestral assessment and evaluation? Why or why not?
3. Is writing considered an important part of the language program of the school you work at? Why or Why not?

Questions about participants´ use of rubrics

4. Do you consider that using rubrics in the assessment of EFL writing is important? Why?
5. Do you use rubrics to give a score to your students? What type of rubric? Why?
6. Which rubric do you prefer to use holistic or analytic? Why?
7. Do you consider that rubric provided improved your scoring of the writing samples?

Questions about the training session

8. Do you consider training is necessary to score writing? Why?
9. Do you consider the training provided may improve your future assessment? Why?
10. Do you consider it necessary to take training to assess students´ written work? Why? Why not?
11. What aspects do you consider can be improved of the training session?

Questions about participants´ experience scoring the sample papers.

12. How did you feel while scoring the papers before taking the training session? What difficulties did you have? Did training help you solve these issues?
13. Do you consider that your scoring of the 10 written samples improved after taking the training? Why or why not? How did it help?

Appendix D Teacher Interview 2 Outline (post to training sessions)

Now that you have experienced two Writing Assessment Training Sessions,

1. Do you continue assessing writing in your EFL classroom?
2. If so, how do you do it?
3. What changes have you implemented in your assessment of writing after the training session?
4. What changes do you intend to implement in your future lessons? Why?
5. Do you now use rubrics to assess your students´ writing? Which? Why?
6. Do you use rubrics to give feedback to your students´ writing? Why?
7. Has your use of rubrics changed after taking the assessment training?
8. How has the training session helped you in your writing assessment practice? Why or Why not?
9. What changes would you make to the training session?
10. How do you feel about writing after taking the training session?
11. How do you feel about writing assessment after taking the training session?
12. Do you have any additional comments?

Appendix E Language Manager Interview 1 Outline

1. What is this EFL program´s teaching goals? And learning goals?
2. Is teaching writing to students a part of those teaching and learning goals? Why or why not?
3. What issues are faced when including the teaching of writing in this EFL program?
4. How do you believe these issues can be solved?
5. Is providing teachers with the appropriate training for the teaching of writing necessary? Why or why not?
6. Did the training provided by the researcher help the management of the language program? If it helped, please explain how.
7. Did the training provided by the researcher help the teachers of the language program better assess their students? If it helped, please explain how.
8. What issues are faced when including writing assessment in this EFL program?
9. How do you believe these issues can be solved?

To conclude, what is your opinion about the following?

Our EFL teachers know that writing is important, they know that teaching writing will result beneficial for students and that by assessing writing in the classroom they give the importance it should have. But why is it that in some institutions the teaching and assessment of writing is not happening in the EFL classroom?

Appendix F Language Manager Interview 2 Outline

Now that your teacher staff have experienced two Writing Assessment Training Sessions,

1. What changes have you observed in their classroom assessment of writing?
2. How do you consider the session has impacted teachers personal assessment practice?
3. Has the training session promoted change in the language program? Why? Or Why not?
4. If so, what type of change?
   If not, do you believe or consider changes in the program in the future?
5. Do you believe it is possible to promote writing assessment by providing training to teachers?
6. How has assessment training changed your personal point of view of writing assessment?
7. The issues you mentioned in our first encounter, has training allowed the program to solve them? Why or Why not?

Appendix G Student Focus Group Protocol 1

Moderator introduces herself and the purpose of the meeting. Moderator allows students to feel comfortable by allowing them to introduce themselves if they wish to do so.

Moderator explains project, purpose of the project and requests permission to record session.

Moderator begins discussion by eliciting the following questions.

a) Do you like to write in English?
b) What do you write?
c) Do you write in your English class? Since when?
d) How did you learn to write in English? What helped you the most?
e) Is your writing considered part of your grade? Why or Why not?
f) Do you agree with having your writing included (or not included) in your English evaluation?
g) If your teacher considers your writing part of your grade, how does she do it?

Appendix H Student Focus Group Protocol 2

Moderator introduces herself and the purpose of the meeting. Moderator allows students to feel comfortable by recalling the information reviewed in the previous session and encouraging reflective discussion.

Moderator describes the information provided by the students in previous session,

Moderator elicits the following questions:

About the student

1. From the beginning of the term to now the end of the term, do you feel about your writing in English?

If improved, in what do you feel it improved?

If not improved, why not?

2. What difficulties were you able to overcome during this term? What difficulties do you still have?

3. What do you believe you still needs to improve?

About the teachers´ assessment

1. How do you feel with the English grade you have obtained for the term? Why?
2. How do you feel about the grade you obtained for writing? Why?
3. From the beginning of the term up until now has the teachers´ assessment of English changed?
   If yes, in what? How did it change?
   Do you agree with the change? Why or why not?
   If no, would you have liked a change? Which change?
4. From the beginning of the term up until now has the teachers´ assessment of writing changed?
   If yes, in what? How did it change?
5. Do you agree with the change? Why or why not?
   If no, would you have liked a change? Which change?

Appendix I Writing Samples and Task Prompt

SAMPLE 1

Task Instructions: People lie everyday and sometimes don't notice it. Others tell "white lies" to avoid hurting people. Have you lied before? Why did you lie? What consequences did it bring? Describe your most "memorable lie" and explain what you said and why you said it. Describe what consequences it had after. Write your description in minimum 120 words, maximum 180 words.

Student´s Written Task:

When I am a childreen I lived in Solina Cruz Oaxaca and one day I got up and my mom went to school I don't want go the school and I said my mom " mom I have stomage ache" and my mom don't belved. but I pretend and my mom preocupete and she back to house and I went to slept I played video game everyday and I ate much food. before my mom said why lie? And I don't said answer my mom angry with me and I said sorry mom but I don't go to school I went to sleep everyday my mom are dsepcionate with me.

SAMPLE 2

Task Instructions: People lie everyday and sometimes don't notice it. Others tell "white lies" to avoid hurting people. Have you lied before? Why did you lie? What consequences did it bring? Describe your most "memorable lie" and explain what you said and why you said it. Describe what consequences it had after. Write your description in minimum 120 words, maximum 180 words**.**

Student´s Written Task:

People lie everybody…
I've lied before, I've lied to many people, but I've lied more to my friends.
The lies I say non-gravity are small and unimportant.
One day I said to my friends I had arrived late to the school because I was sick but the truth is I stay sleep in my home and made me late.
I lied because I thought that it would be shameful to arrive late to school and maybe the teachers wouldn't live me enter to the classroom.
The consequenses were any bad, I couldn't present the exam of english and not give my homework to my teacher of english.
I think if I had gone to school the consequences would have been minor.
Now I try not to lie because it is not good for anyone, and ther is to be an honest person.

SAMPLE 3

Task Instructions: People lie everyday and sometimes don't notice it. Others tell "white lies" to avoid hurting people. Have you lied before? Why did you lie? What consequences did it bring? Describe your most "memorable lie" and explain what you said and why you

said it. Describe what consequences it had after. Write your description in minimum 120 words, maximum 180 words**.**

Student´s Written Task:

Yes, I have lied before. I think I did because I did not want to get in trouble, I was just a kid. Lying brought me a little consequences, nothing very important. My memorable lie was in the 90's with my cousin. We were just a two little girls making mischief. So one day we were playing in the bathroom. In our game we set fire with candles and paper. In a moment everything was out of control so we tried off the fire. We ran from there. An our later our parents asked for the situation. We said we didn't know anything. 100% denegation. That afternoon my grandmother said to our parents that she saw us out of the bathroom. Obviusly they believe to my grandmother. The consequence was that they put as a punishment for a week. Very sad.

SAMPLE 4

Task Instructions: People lie every day and sometimes don't notice it. Others tell "white lies" to avoid hurting people. Have you lied before? Why did you lie? What consequences did it bring? Describe your most "memorable lie" and explain what you said and why you said it. Describe what consequences it had after. Write your description in minimum 120 words, maximum 180 words**.**

Student´s Written Task:

Well this happens some years ago, I lied because in that moment was necesary to not be discovered, but with the pas of time this lie was discovered and my mother said me "This can't repeat", I did it because i was learn to drive standar and I took the car and I drive arround the my neighborhood.
My mom worked all mornings so when she and my father going to the work I take the car's keys and I drive the car. Oviusly in some days I could drive efficiently and this lie wasn't necessary and everything come back to reality.
The consecuenses of this lie, I think was satisfactory because nobody was injured and I learn to drive standar in some days and my parents didn't notice about it so I wasn't punished and all is well.
I think, I do it again

SAMPLE 5

Task Instructions: People lie everyday and sometimes don't notice it. Others tell "white lies" to avoid hurting people. Have you lied before? Why did you lie? What consequences did it bring? Describe your most "memorable lie" and explain what you said and why you said it. Describe what consequences it had after. Write your description in minimum 120 words, maximum 180 words**.**

Student´s Written Task:

Have you lied before? Yes

My most memorable lie, was when I was a child I till remember, when I caused an accident with a game pyrotechnic, burning a lonely place I was very scared and I had to lie, to avoid being scolded for my father.

Why did you lied? because I thought that my father me would punish. I remember that my father question me "what have happened here?" and my answer was I do not know, I just saw the flames.

What consequences did it bring?

The consequences of this was, when my father discovered, my lie. He caught my attention, warning me do not lie

Appendix J Analytic Rubric

Score	Content	Organization	Use of Language	Use of Vocabulary	Mechanics and Spelling.
5	Text shows knowledge of the topic and gives details or examples to support main ideas. Text fully corresponds to task requirements. Communication is effective.	Organizational skills are present in the text making flow and coherence of ideas smooth. Main ideas and structure of text are easily found and logically sequenced.	Text makes use and maintains use of complex language structures effectively. There are no errors of idioms, collocations and grammar in general. Facility in use of language is apparent.	Demonstrates sophisticated and broad use of vocabulary. Effective and appropriate use of idiomatic expressions and colloquialisms; shows awareness the connotations and their meaning.	Writing presents mastery of punctuation and spelling conventions. Errors of capitalization, paragraphing and typos are not found
4	Task is answered in its majority but information may be redundant or unnecessary. Some detail is given. Sufficient development of main ideas. Some gaps may be found among information.	Adequately organized with the use of organizational patterns, and connectors but sequencing of information is incomplete. Connection of main ideas may be lost but meaning is still understood.	Grammatical accuracy consistently maintained; Few errors of idioms, collocations and grammar in general. Complex sentences present minor errors.	Demonstrates sophisticated use of vocabulary. Good command of Idiomatic expressions and colloquialisms. Minor vocabulary use errors but not significant.	Writing presents occasional errors of punctuation and spelling conventions. Errors of capitalization, paragraphing and typos are occasionally found.
3	Task is addressed adequately but information may be missing. Some details are used to support the main idea. Shows some knowledge of the main topic and limited development of main ideas.	Some organizational skills are present. Use of cohesive devices makes text clear and understood. Occasional deficiencies can lead to “jumpiness” among information.	Some grammatical “slips” may be found. Grammatical errors such as verb tense, verb agreement, number, word order, articles, pronouns, and prepositions are found but they do not lead to misunderstanding. Context given in text allows for interpretation of meaning.	Vocabulary accuracy is high though occasional errors may be found. Adequate and appropriate word/idiom choice and use. Some incorrect word choice does occur without impeding communication.	Writing presents few errors of punctuation and spelling conventions. Few errors of capitalization, paragraphing and typos are found.
2	Task reveals little relevance to the topic. Major gaps in information are found and insufficient details to support main ideas are given. Inappropriate information. Pointless repetition of information.	Small pieces of text are linked with basic connectors. Unsatisfactory cohesion may cause most but not all, of the information to seem sloppy and non-fluent.	Frequent grammatical inaccuracies found. Frequent and basic errors of tense, agreement, number, word order, articles, pronouns, and prepositions are found. Understanding of ideas is seldom confusing.	Sufficient control of elementary vocabulary to express basic ideas. Repetition of vocabulary is frequent. Frequent misuse of word form use, word/idiom choice and use making communication confusing.	Writing presents frequent errors of punctuation and spelling conventions. Errors of capitalization, paragraphing and typos are frequently found. Meaning may be confusing.
1	Task presents limited relevance to main topic. Inadequate development of topic. Details are not given.	Groups of words connected with simple connectors such as “and”, “but” or “because”. Cohesion is almost absent. Connection among ideas is difficult to find making information confusing or misleading.	Almost all or most of the basic grammatical constructions are inaccurate. Major issues in simple sentences. Errors of negation, agreement, number, word order, articles, pronouns, prepositions frequently found. Understanding of information difficult.	Text has little knowledge of English vocabulary, idioms and word forms. Has sufficient for coping with simple survival needs. Information is basically translated. Inappropriate choice of word form.	Almost all of the spelling is inaccurate and ignorance of punctuation conventions among text is found. Text is dominated by capitalization, paragraphing and typo mistakes. Meaning is obscured.
0	Task does not reveal development topic. Totally inadequate answer to task. No details are given. Content insufficient to assess.	Cohesion is totally absent. Writing is fragmented making communication impossible to obtain. Lack of structure in information leads to absence of organization. Content insufficient to assess.	All language use is inaccurate. Meaning obscured. Content insufficient to assess.	No apparent vocabulary use and vocabulary comprehension is present in text. Content insufficient to assess.	All of the spelling is inaccurate and ignorance of punctuation conventions among text is found. Text is dominated by capitalization, paragraphing and typo mistakes. Meaning is obscured. Content insufficient to assess.

Appendix K Holistic Rubric

Score	Description: A written text may include all or some of the following characteristics.
5	Text is clear and ideas smoothly flow from one section to another. Writing is a complex text about own experiences providing more than enough details. Style of text is appropriate and sequence of information is logical helping the reader understand and find main ideas. Writer focuses on producing a text that corresponds to task required.
4	Text is clear and/or is a detailed description of familiar topics that point out important issues by providing personal points of view or experiences. Reasons and important examples are given to support points of view. A structure in the text is found and concludes with a clear and understandable conclusion. Most of the text corresponds with the required task.
3	In text points of view are provided occasionally supported by reasons or examples. A relationship can be found among ideas and organization may be adequate. Text corresponds with task required but gaps may be found. Rarely, meaning in communication may be difficult to find.
2	Very short text that describes basic experiences, feelings, reactions and/or a sequence of linear events in simple connected sentences. Reasons, examples or points of view may be inadequate or not enough to support main ideas. Organization may be inadequate. Frequently main ideas are obscured or deviate from the main task.
1	Text is a series of simple phrases and sentences linked with simple connectors like **'and'**, **'but'** and **'because'**. Sentences are disorganized and lack meaning. Details or points of view are not given. Text presents problems of focus on task required.
0	Simple isolated phrases and sentences are given with no structure or organization. Meaning is obscured and ideas do not correspond to task required.

Appendix L On-Line Post-Training Questionnaire Protocol

This questionnaire is part of a research project that seeks to analyse the assessment strategies of EFL writing in distinct universities of Ciudad Victoria Tamaulipas Mexico. This instrument has the purpose of knowing your opinion in relation to the training session to which you assisted and to know more about its effectiveness. The information you share on this questionnaire is anonymous and confidential. It will only be used for research purposes.

Date on which you took training: _____
Place at which you took training: _____
Participant ID: _____
Instructions: Please be so kind and honestly give your opinion in relation to the following statements by ticking the box that corresponds to your opinion. Consider that the numbers have the following meanings:

1: Strongly Agree **4. Disagree**
2. Agree **5. Strongly Disagree**
3. Neither Agree or Disagree

Statement	**1**	**2**	**3**	**4**	**5**
1. The information and practice shared during the training session was **clear and understandable.**	☐	☐	☐	☐	☐
2. The information and practice shared during the training session is **practical** for my future evaluation of students´ writing.	☐	☐	☐	☐	☐
3. The information and practice shared during the training session is **useful** for my future evaluation of students´ writing.	☐	☐	☐	☐	☐
4. After taking the training session, I consider that **my use** of rubrics has become more **efficient.**	☐	☐	☐	☐	☐
5. After taking the training session, I consider that **my use** of rubrics has become **easier.**	☐	☐	☐	☐	☐
6. After taking the training session, I have decided to use an evaluation tool such as a rubric to assess my students´ writing skills.	☐	☐	☐	☐	☐
7. The rubrics provided by the researcher/trainer will be useful for my future evaluations of writing.	☐	☐	☐	☐	☐
8. After the training session, the scoring of the writing samples provided by the researcher was **easier**.	☐	☐	☐	☐	☐
9. After the training session, the scoring of the writing samples provided by the researcher was more **efficient**.	☐	☐	☐	☐	☐

Instructions: Please read the following statements and choose the answer (s) that best corresponds to your opinion. Where necessary, explain your choice.
10. In your opinion, what aspects should an ideal scoring training session include? Choose the options you consider necessary from below. More than one option may be chosen.
☐ Theoretical background to the evaluation of writing.

☐ Discussion of the distinct types of rubrics and their use.
☐ Group/Individual analysis of writing samples and their scores.
☐ Group discussions of writing samples and their scores.
☐ Group/individual scoring practice of writing samples.
☐ Other, please explain_____

11. I consider **my understanding and use of scoring rubrics** has changed after participating in the training session. If your answer is **yes**, please describe in what way your understanding of rubrics has changed. If your answer is **no**, please explain why not.

☐ Yes ☐ No_____

12. After participating in the training session, I prefer using a _____ rubric to score my students´ writing. Choose one option from below and further explain your choice.

☐ Holistic Rubric (Giving a single general score based on first impression.)
☐ Analytic Rubric (Giving several scores for distinct aspects.)
☐ Both Analytic and Holistic.
☐ Other: _____

13. I consider my general writing assessment practice has improved or will improve after participating in the training session. If your answer is **yes**, please describe in what way your assessment practiced has improved. If your answer is **No**, please explain why you consider your assessment activities have not improved.
☐ Yes ☐
No_____

14. Are there any further comments you would like to provide regarding the training session or your use of rubrics to evaluate writing?

_____
_____

Thank you very much for your information!

UNIVERSITY OF Southampton

Appendix M Participant Information Sheet

December 2013, Version 1.

Study Title: The Impact of EFL University Professors' Assessment Training on Classroom Writing Assessment: Practice and Perceptions

Researcher: Elsa Fernanda Gonzalez **Ethics number: 8729**

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?

My name is Elsa Fernanda Gonzalez and I am MPhil/PhD student at the University of Southampton UK. I am currently working with my research project which has the purpose of exploring the impact that teacher training sessions can have on the assessment/evaluation of EFL writing tasks. I am interested in finding out if teacher-training sessions can be beneficial for the scoring of student writing papers and can lower the variability among scores. If you decide to take part in the study your participation would consist of:

1. Answering a background questionnaire in which you will be asked information about your teaching experiences and your perceptions of writing assessment.

2. Scoring 5 writing samples of intermediate EFL students. Samples will be provided to you in printed form by the researcher.

3. Taking part in to training sessions that will last approximately 2.5 hours and will be delivered on three different dates so you can decide which best suits your needs.

4. Scoring once again the 5 written samples that were evaluated before the training.

5. Answering a questionnaire in which you have the opportunity to express your insights in relation to the training session.

6. Take part in two face-to-face interview in which questions about EFL writing assessment and training sessions will be asked.

7. Provide a sample of your students writing done throughout the year.

8. Facilitate two interview sessions with five to ten of your students.

Why have I been chosen?

This study focuses on the scoring of English as a Foreign Language (EFL) writing samples. It seeks to describe the impact that EFL teacher training can have on scoring variability. Therefore EFL teachers that work at public higher education institutions are eligible for this study.

Professional or academic experience is not a determinant factor for your eligibility, as long as you are an in-service teacher working at a public university in Cd. Victoria Tamaulipas Mexico.

What will happen to me if I take part?

If you decide to take part in the study you will be asked to participate in the previously mentioned activities in the following order: 1) answer the background questionnaire to provide information about your teaching background; 2) score 10 writing samples on paper independently prior to the training session; 3) take part in the training session for approximately of 2.5-3 hours. The researcher will deliver sessions on three different dates that will be communicated to you as soon as the place is confirmed. It is only necessary for you to attend ONE session; 4) score the same 10 written samples once again after the training so the researcher can analyse the changes that the training had on the scoring. This scoring process will also be carried out independently and you will be given from 2-3 weeks to score papers; 5) once the training sessions finalized and you have finished rescoring your 10 writing samples, you will be asked to fill in a questionnaire in which you will be asked about your perceptions towards training sessions and how useful you perceived them; finally 6) those who further agree to do so, will be interviewed face-to-face by the researcher with the purpose of obtaining further explanations of your perceptions of writing assessment and the training activities in which you took part.

Are there any benefits in my taking part?

By participating in this research project, you could benefit from:

1. Updating your English teaching skills with a free workshop delivered to you in the comfort of your workplace and obtaining an assistance diploma provided by the institution at which the training would take place.

2. Sharing your experiences and hearing those of other fellow teachers that could benefit your own experience in the evaluation of writing.

3. Having the personal satisfaction of being part of a research project that seeks to improve English language teaching in our city.

Are there any risks involved?

By taking part in this study you may:

1. Feel overwhelmed with the scoring of student papers that needs to be done for this study and the responsibilities that need to be fulfilled at your job or working place.
2. Feel that the workload at your own job and that of attending the workshop may be too much. You may also be at risk of not being able to attend any of the training sessions because your own time schedule may not allow it.
3. Feel uninterested or tired during workshop and fail to pay attention. However, training sessions will be delivered as dynamic as possible and will allow for much participation to avoid these feelings.
4. Feel confusion or misunderstanding of what you need to do with the writing samples. If any doubts or inquires arise you are welcome to contact the researcher at any time for further explanation.

Will my participation be confidential?

The information you provide in the different questionnaires as well as your scores will be anonymous and confidential in compliance with the Data Protection Act 1998 and the University's Data Protection Policy and Guidelines. Although your questionnaires and interviews will not require you to state your name or any other personal information, scores obtained from the study will be coded with the purpose of relating the first batch of writing samples that you scored with the second batch of papers. Your identity will not be traced in any way (linked anonymity). Your work and scores will be stored by the researcher on a password-protected computer and shared only with the supervisor of this study for analysis purposes.

What happens if I change my mind?

If you happen to change your mind and decide to withdraw your participation, your legal rights or workplace situation will not be affected. You may drop out of the study any time you feel it is convenient.

What happens if something goes wrong?

If at any time during the study you feel concerned or wish to file a complaint to someone that is not the researcher you may do so by contacting the Chair of the Faculty Ethics Committee of the University of Southampton Prof Chris Janaway by telephone to the number 023 80593424 or by sending an email to c.janaway@soton.ac.uk.

Where can I get more information?

For further information regarding your participation or the information outlined in this Participant Information Sheet please feel free to contact me by dialing my **cell phone (834) 1160176 or by sending an email to e.fernandagonzalez@gmail.com.**

Appendix N Informed Consent

CONSENT FORM *(FACE TO FACE:* December 2013, Version 1)

Study title: **The Impact of EFL University Professors' Assessment Training on Classroom Writing Assessment: Practice and Perceptions**

Researcher name: Elsa Fernanda Gonzalez
Staff/Student number: efg1y12

ERGO reference number: 8729

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet (insert date /version no. of participant information sheet) 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……